



The Developmental Trajectory of Second Language Listening

Errors, Processes and Change in the Understanding of Beginner Learners of French

(Elizabeth) Kedi Simpson

Thesis submitted for the degree of

DPhil in Education

Michaelmas Term 2024

Department of Education,

Kellogg College,

University of Oxford



Graffiti in the rue Lhomond, 75005 Paris

Charles-François Lhomond was an educator and grammarian and author of
Les Elemens de la grammaire françoise (1771)

Original quote by Henri Matisse: *l'exactitude n'est pas la vérité*
(‘precision is not truth’)

Abstract

This thesis aims to provide an alternative view on how students learn to listen. Most research in this area takes a cross-sectional approach in order to demonstrate factors which might influence listening proficiency at a given time point, but this ‘snapshot’ cannot tell us what might have come before, or what might come after. This means that little is known about how a group’s or an individual’s listening proficiency might develop over time, and the relative importance of different factors at different places on their journey.

To address this problem, the present study tracked the development of L2 listening of learners of French in an English secondary school over a period of three years, from when they began secondary school, aged 11, until the end of statutory language tuition aged 14. The research takes a detailed perspective on the developmental trajectory of L2 listening.

Using a Complex Dynamic Systems approach, data were collected at twenty time points over three academic years to track the development of participants’ listening proficiency by means of a paused translation protocol. In addition, subsystems that contribute to listening success were measured: lexical and syntactic knowledge, strategy use, motivation and working memory. Interviews of four case study participants took place at each time point and included stimulated recall. The findings shed light on both intra- and inter-individual variation during the language acquisition process to allow the exploration of complex dynamic systems at from moment-to-moment, and over a three-year period both at a whole cohort level, and through the lens of four case studies.

The thesis begins by analysing errors made by the cohort and presents them within the context of John Field's model of listening (Field, 2008b, 2013), in order to ask what the errors reveal about the participants' listening processes, uncovering the complex dynamic systems apparent during the moment-to-moment experience of listening, as well as how these errors evolve over the three years in question. Findings are to an extent consistent with the literature – for example that vocabulary knowledge plays a larger role than grammatical knowledge, and that the L1 is an ever-present influence – but also reinforce the significance of the modality of input in the comprehension process. Findings also suggest other new factors that might be at play, such as the influence of cognates and false friends when struggling to interpret aural input, the challenge of contrastive voice onset timing between French and English, and the salience of consonants over vowels.

The thesis then zooms out to explore progress in listening proficiency, conceptualised as a complex dynamic system in which linguistic subsystems (vocabulary and grammatical knowledge) and metacognitive subsystems (strategic behaviour, self-efficacy, and anxiety) interact over time. Using a model of moving correlations, descriptive statistics probe the patterns of interaction between the subsystems and a picture is built of the range of variability which might occur and how the different subsystems vary in importance over time. For the whole cohort, it shows that vocabulary knowledge has a consistently moderate correlation with listening proficiency, while grammatical knowledge has a consistently weak correlation. There are periods where the correlations with listening proficiency and either of these subsystems appear to grow together, and other periods where one correlation increases while the other decreases. As far as metacognitive systems are concerned, the four elements seem to split into two distinct pairs: the moving correlations between both Problem-Solving and Mental Translation and

listening mirror each other, as do those between both Person Knowledge and Directed Attention and listening.

Finally, four case study participants are examined, allowing the boundaries of the last complex dynamic system to include contextual data about the students' classroom experience as well as their broader lives. This permits a demonstration of the extent to which to the above questions vary between individuals, and how individual patterns compare to those seen at group level, as well as additional influences on students' learning, such as additional home languages, specific thinking patterns, or special educational needs.

Implications are far reaching and are divided into four main categories. The theoretical implications return to the theoretical frameworks on which the present study is based and probe them in the light of the findings. Scholarly implications discuss further directions for research into L2 listening, suggesting new hypotheses to be tested. Methodologies are also addressed given the scope of the present study. The pedagogical implications examine L2 listening pedagogy in terms of teaching (particularly teaching teenagers in input-poor environments), and resource creation.

Acknowledgements

For Jo, my biggest cheerleader – and frequently cook, cleaner and patient observer while I sit for yet another evening and write, write, write.

Thanks also to...

My supervisors Professor Heath Rose and Dr Robert Woore, who have been supportive and patient even when I have not been, and to Professor Wander Lowie at the University of Groningen who pointed me in the right direction for CDST on many occasions, and to Professor Mike Horton at Leeds whose Rasch course was the saving of me. Furthermore, of course to my wonderfully supportive examiners Dr Laura Molway and Dr Sal Consoli.

Jake, Flo and Isaac, my children when this started, now my adults. Flo created the illustration in Figure 5.

My parents whose financial shove in the right direction is very much appreciated.

Other friends and family who have checked in, asked about progress, shown interest:

Garry and John, Sam and Heather, Tikki and Rebecca, fellow DPhillers Stephen, Ikuya, Darsh, Kari, and Heath's little gang of supervisees, the 'academic superchums' What's App group.

My colleagues at school who tolerated my being part-time, who booked computer rooms for me, or let my interviewees out of their lessons for yet another 'quick chat'. Thanks also to the full cohort of 80 students who patiently did test battery after test battery, and of course to Helen, Theo, Ryan and Fran, my fabulous case studies.

Contents

1. Introduction	18
1.1. Research context	18
1.2. Research objectives, questions, scope	18
1.3. Outline of the study	20
1.4. A personal note	22
2. Literature Review	24
2.1. Introduction	24
2.2. Theoretical frameworks.....	24
2.2.1. Theoretical framework for the process of listening	24
2.2.2. The Fuzzy Lexical Representation hypothesis.....	26
2.2.3. Listening to learn, or learning to listen?.....	28
2.3. Measuring listening proficiency.....	31
2.3.1. The Comprehension Approach: the product of L2 listening.....	31
2.3.2. Error analysis in listening.....	33
2.4. Listening and subsystems.....	35
2.4.1. Listening and vocabulary knowledge.....	36
2.4.2. Listening and segmentation skills	37
2.4.3. Listening and grammatical knowledge	38
2.4.4. Listening and metacognition	39
2.5. Longitudinal research in L2 listening.....	44
2.5.1. Strategy use: the process of L2 listening?.....	44
2.5.2. Multi-wave longitudinal research: introducing CDST research	48
2.6. Longitudinal research within a CDST framework	51
2.6.1. The advantages of a CDST approach.....	53
2.6.2. A match with theoretical frameworks of listening?.....	53
2.6.3. Precursors and dependents	54
2.6.4. Complexity as messy yet valid.....	56
2.6.5. The challenges associated with CDST research.....	57
2.6.6. Summary	58
2.7. Research questions	58
3. Methodology	60
3.1. Theoretical and ontological orientation.....	61
3.1.1. Positionality.....	62
3.2. Overview of research design	65
3.2.1. Aims and research questions	65
3.2.2. Population	66
3.2.3. Institutional context.....	66
3.2.4. Participants.....	67
3.2.5. Testing schedule	69
3.2.6. Locations and equipment	70
3.3. Methodological challenges and possible solutions	70
3.3.1. Operationalising understanding spoken language.....	70
3.3.2. Longitudinal challenges and solutions	72
3.3.3. CDST and the challenge of the boundary	75
3.4. Piloting instruments.....	75
3.4.1. Choice of listening passages	76
3.4.2. Choice of response to listening passage.....	76
3.4.3. Piloting the listening passages	77

3.4.4. Choice of scoring the paused translation passages	80
3.4.5. Measuring vocabulary knowledge	83
3.4.6. Piloting the AuralLex	83
3.4.7. Measuring grammatical knowledge	89
3.4.8. Piloting the AuraGram	89
3.4.9. Connection between vocabulary and grammar	95
3.4.10. Measuring strategy use.....	95
3.5. Data collection procedure.....	96
3.5.1. Listening proficiency	96
3.5.2. Vocabulary knowledge	98
3.5.3. Grammatical knowledge	99
3.5.4. Metacognition	100
3.5.5. Semi-structured interviews.....	101
3.6. Data analysis procedure	102
3.6.1. RQ1: listening error analysis.....	102
3.6.2. RQ2: relationships between subsystems	104
3.6.3. RQ3: case studies	108
3.7. Ethical considerations	110
3.7.1. Minimising harm.....	110
3.7.2. Position as a teacher-researcher	111
3.7.3. Data and record-keeping	111
3.8. Problems arising	112
3.8.1. The impact of the Covid-19 pandemic.....	112
3.8.2. Test fatigue and attrition.....	114
3.9. Summary of chapter	115
4. Findings 1: errors and listening processes	116
4.1. Introduction to the chapter	116
4.1.1. Explanation of chapter structure	117
4.1.2. The influence of covid lockdowns	122
4.1.3. Cautious, ambitious, and reckless test-takers.....	123
4.1.4. A deficit model of error analysis	125
4.1.5. Codebook summary	125
4.2. Phonologically driven whole-track segmentation difficulties.....	127
4.2.1. Missed sections, and changes over time with missed sections	128
4.3. Phonologically driven vocabulary errors	130
4.3.1. Word boundary difficulties.....	131
4.3.2. Word boundary difficulties over time	142
4.3.3. Further lexico-phonological errors.....	146
4.3.4. Active knowledge is no guarantee of passive recognition	154
4.3.5. Changes over time in phonological errors	155
4.3.6. Summary of lexico-phonological findings.....	158
4.4. Textually driven vocabulary errors.....	158
4.4.1. Textual false friends	159
4.4.2. Lexical sets.....	161
4.4.3. Teaching-dependent errors	162
4.4.4. Summary of textually driven vocabulary errors.....	165
4.5. Grammatical errors.....	165
4.5.1. Content and function words	165
4.5.2. Difficulty in recognising morphological features	169
4.5.3. Summary of grammatical errors.....	172

4.6. Overarching conclusion of chapter	172
5. Findings 2: relationships between subsystems.....	174
5.1. Listening proficiency and its variability.....	175
5.1.1. Whole sample listening scores	175
5.1.2. Variability in paused translation scores: whole sample.....	176
5.2. Linguistic subsystems and their variability.....	179
5.2.1. AuralLex scores.....	179
5.2.2. AuraGram scores.....	183
5.2.3. Developing relationship between linguistic subsystems.....	186
5.2.4. Linguistic subsystems: concluding remarks.....	190
5.3. Metacognitive subsystems, their development and relationships	191
5.3.1. Metacognitive scores.....	192
5.3.2. Developing relationships between metacognitive subsystems	194
5.3.3. Metacognitive subsystems: concluding remarks.....	197
5.4. Summary of findings 2.....	197
6. Findings 3: Helen	199
6.1. Helen: Background information.....	199
6.2. Helen's listening proficiency and its variability	201
6.2.1. Variability in listening proficiency scores.....	201
6.3. Helen's linguistic subsystems, their development and relationships	203
6.3.1. AuralLex scores.....	203
6.3.2. AuraGram scores.....	205
6.3.3. Developing relationship between linguistic subsystems.....	207
6.4. Helen's errors and listening processes	208
6.4.1. Perception of speed	209
6.4.2. Favouring consonants over vowels	209
6.4.3. Sound and sense work together.....	211
6.4.4. Content and function words	212
6.4.5. Morphological features	214
6.5. Summary of findings for the linguistic systems.....	215
6.6. Helen's metacognitive subsystems, their development and relationships	217
6.6.1. Metacognitive scores.....	217
6.6.2. Developing relationships between metacognitive scores.....	218
6.7. Helen's strategic journey.....	220
6.7.1. Phase 1: An anxious mental translator determined to succeed	220
6.7.2. Phase 2: Helen as a 'self-subtitler'.....	222
6.7.3. Phase 3: Growing confidence and proficiency.....	224
6.7.4. Metacognitive subsystems: Concluding remarks.....	227
6.8. Helen's complex dynamics: conclusion.....	228
7. Findings 4: Theo.....	229
7.1. Theo: Background information	229
7.2. Theo's listening proficiency and its variability.....	231
7.2.1. Variability in listening proficiency scores.....	232
7.3. Theo's linguistic subsystems, and their development and relationships.....	234
7.3.1. AuralLex scores.....	234
7.3.2. AuraGram scores.....	236
7.3.3. Developing relationship between linguistic subsystems.....	238
7.4. Theo's errors and listening processes.....	240
7.4.1. A listener unfazed by the French speech stream	241
7.4.2. Puzzling it out with vocabulary at the core.....	241

7.4.3. A preoccupation with verbs yet overlooking morphological features.....	244
7.5. Summary of findings for the linguistic systems.....	247
7.6. Theo's metacognitive subsystems, their development and relationships.....	247
7.6.1. Metacognitive scores.....	247
7.6.2. Developing relationships between metacognitive scores.....	249
7.7. Theo's strategic journey	253
7.7.1. A listener who relishes challenge and expects to do well	253
7.7.2. A perfectionist learner relieved by the lack of control in listening	254
7.7.3. Translating word by word or chunk by chunk	255
7.7.4. Theo's metacognitive systems: concluding remarks.....	258
7.8. Theo's complex dynamics: conclusion	258
8. Findings 5: Fran	260
8.1. Fran: Background information.....	260
8.2. Fran's listening proficiency and its variability.....	263
8.3. Fran's linguistic subsystems, their development and variability	266
8.3.1. AuralLex scores.....	266
8.3.2. AuraGram scores	267
8.3.3. Developing relationship between linguistic subsystems.....	268
8.4. Fran's errors and listening processes	270
8.4.1. False friends that respect word boundaries	270
8.4.2. Listening with Italian ears: a different set of challenges.....	273
8.4.3. Plurals and tenses: morphological features are not detected.....	275
8.4.4. Common sense fights with perception	277
8.5. Summary of findings for the linguistic systems.....	277
8.6. Fran's metacognitive subsystems, their development and variability.....	278
8.6.1. Metacognitive scores.....	278
8.6.2. Developing relationships between metacognitive subsystems	280
8.7. Fran's strategic journey	282
8.7.1. Breathing deeply, eyes closed to keep calm.....	283
8.7.2. A desire to please and a fear of judgement.....	284
8.7.3. A mature and big-hearted attitude to learning.....	285
8.7.4. 'You press a button in your head': Mental Translation / Problem-Solving ..	286
8.7.5. Fran's metacognitive subsystems: concluding remarks.....	289
8.8. Fran's complex dynamics: Conclusion	289
9. Findings 6: Ryan	290
9.1. Ryan: Background information.....	290
9.2. Ryan's listening proficiency and its variability.....	293
9.2.1. Variability in listening proficiency scores:.....	293
9.3. Ryan's linguistic subsystems, their development, and relationships	295
9.3.1. AuralLex scores.....	295
9.3.2. AuraGram scores	297
9.3.3. Developing relationships between linguistic subsystems	299
9.4. Ryan's listening errors and processes	300
9.4.1. Understanding dominated by cognates	301
9.4.2. Words but rarely sentences.....	302
9.5. Summary of findings for Ryan's linguistic systems	304
9.6. Ryan's metacognitive subsystems, their development and variability.....	304
9.6.1. Metacognitive scores.....	304
9.6.2. Developing relationships between metacognitive subsystems	307
9.7. Ryan's strategic journey	309

9.7.1. ‘I didn’t mean to hit Sir’: An impulsive sense of fun	309
9.7.2. An inconsistent message	311
9.7.3. Times 1-6: ‘I don’t know what it means, but I kind of wanted to do it’	312
9.7.4. Times 7-11: ‘I’ve always got them right’ but ‘it just poofs out of my mind’	315
9.7.5. Times 12 to 17: ‘I searched up how to be a better rememberer’	317
9.7.6. Times 18 to 20: ‘I knew it was wrong, but I wanted to put something’	318
9.8. Ryan’s metacognitive subsystems: Concluding remarks	319
9.9. Ryan’s complex dynamics: Conclusion	319
10. Discussion and analysis.....	321
10.1. Introduction	321
10.2. Group level subsystems over time	322
10.2.1. Linguistic subsystems at group level	322
10.2.2. Relationship between listening proficiency and vocabulary.....	323
10.2.3. Relationship between listening proficiency and grammar	331
10.2.4. Metacognitive systems at group level	334
10.3. Case studies over time.....	336
10.3.1. The ergodicity problem	337
10.3.2. Helen as ‘self-subtitled’ (reprise).....	341
10.3.3. Helen and Theo: Alternating between vocabulary and grammar.....	343
10.3.4. Theo and Fran: Home languages, affect, and life capital.....	344
10.3.5. Remote learning and its implications.....	345
10.3.6. Ryan: Special needs, covid, and a widening ‘gap’	347
10.4. Discussion: Closing remarks.....	350
11. Implications	351
11.1. Introduction	351
11.2. Theoretical implications	351
11.2.1. Field’s listening model and its subsystems	353
11.2.2. Listening to learn: Noticing grammar	355
11.2.3. Receptive interlanguage	358
11.2.4. The ergodicity problem (reprise).....	358
11.2.5. Theoretical implications: Mini conclusion.....	359
11.3. Scholarly implications	359
11.4. Methodological implications.....	361
11.4.1. Longitudinal approach	361
11.4.2. Ethical implications of being a teacher-researcher	362
11.4.3. Complex Dynamic Systems Theory.....	363
11.4.4. Instruments.....	363
11.4.5. Participants.....	366
11.5. Pedagogical implications.....	368
11.5.1. Listening to learn (reprise).....	368
11.5.2. Learning to listen: Individual differences in the classroom	369
11.5.3. Learning to listen: Teaching students with additional needs.....	370
11.5.4. Learning to listen: Awareness of the influence of L1.....	371
11.5.5. Learning to listen: Implications for resource creation	371
11.6. Policy implications	375
11.7. Conclusion.....	376
12. Limitations	377
12.1. Methodological limitations	377
12.1.1. Panel conditioning.....	377
12.1.2. Access to equipment in school	378

12.1.3. Linguistic tests	378
12.1.4. Use of a single school	379
12.1.5. Use of my own school.....	380
12.2. Analytical limitations	381
12.2.1. Conceiving of the model	381
12.2.2. Correlation windows	382
13. Conclusion.....	383
14. References	386
15. Appendices	406
15.1. Piloting	406
15.1.1. Metacognitive Awareness Listening Questionnaire	406
15.1.2. Additional question	407
15.1.3. Working memory tests	407
15.2. Which paused translation test when	408
15.3. Full transcripts of all listening tests, plus English translations	408
15.4. Full list of AuralLex tests	415
15.5. Full list of AuraGram tests	417
15.6. Excerpts from reflexive diary for coding the paused translation	421
15.7. Examples of paused translation mark scheme.....	424
15.8. Rasch model underpinning statistics	427
15.9. Semi-structured interview protocol – sample from T11	429
15.9.1. Typical interview transcript.....	430
15.10. Codebook summary.....	443
15.11. Screenshots demonstrating the evolution of coding.....	446
15.11.1. Codebook 3.1.23	446
15.11.2. Codebook 2.4.23	446
15.11.3. Codebook 30.5.23	447
15.12. Ethics paperwork.....	448
15.13. Criteria for exclusion from study	449
15.14. Summary of test difficulty levels	450

Tables

Table 1. Summary of recent longitudinal research into L2 listening	46
Table 2: Summary of research questions, sample, instruments, and data analysis	65
Table 3: National context of study setting	67
Table 4: Languages to which participants are exposed outside of school.....	68
Table 5: Details of case studies	68
Table 6. Test dates and term dates for first year	69
Table 7: Paused translation pilot: spreadsheet to find appropriate passages.....	77
Table 8: Developing the 'Guiding Principles' of marking the paused translations.....	82
Table 9: Example of final mark scheme for paused translation	82
Table 10: Aurallex Pilot: Multivariate tests to establish similarities between tests	85
Table 11: AuralLex Pilot: Descriptive statistics for remaining four Aurallex tests.....	85
Table 12: Aurallex Pilot: Testing the assumption of sphericity for Aurallex 3,5,6,7	86
Table 13: Aurallex pilot: Pairwise comparisons - differences between tests	86
Table 14: Aurallex Pilot: Descriptive statistics to investigate floor and ceiling effects....	88
Table 15: Aurallex Pilot: difference between year groups	88
Table 16: Auragram Pilot: location of differences between Auragram tests.....	93
Table 17: no remaining differences between Auragram tests.....	93
Table 18: Descriptive statistics for Auragram tests, by year group.....	94
Table 19: Data collection procedure.....	96
Table 20: Testing protocol.....	97
Table 21: Which AuralLex test when?	98
Table 22: Location of teaching and data setting throughout the data collection period..	112
Table 23: Additional details on remote data collection	113
Table 24: More detailed chapter structure breakdown	118
Table 25: Average scores in paused translation, and percentage of attempts.....	122
Table 26: Paused translation errors: NVivo codebook summary	126
Table 27: Errors coded to segmentation	127
Table 28: All examples of French-derived missegmentations	135
Table 29: Word boundary errors and attempts at translation for test C48.....	143
Table 30: Principal segmentation errors recorded from D25	143
Table 31: Tests which indicate a decrease in word boundary errors	144
Table 32: Tests which indicate no or minimal change in word boundary errors.....	145
Table 33: Instances of syllable metathesis across the dataset	151
Table 34: All examples within the dataset of a double translation.....	153
Table 35: Total number of references of the lexical set codes	162
Table 36: Incorrect 'I like's in repeated tests	164
Table 37: Codebook summary words missing vs words wrong.....	166
Table 38: Missing and wrong content and function words	166
Table 39: Codebook breakdown of grammatical errors	169
Table 40: Descriptive statistics for listening proficiency	176
Table 41: Descriptive statistics for AuralLex scores	181
Table 42: AuraGram descriptive statistics for whole cohort.....	184
Table 43: Moving correlations for linguistic subsystems for the whole cohort.....	189
Table 44: Relationships between listening and metacognitive subsystems	195
Table 45: Helen's listening proficiency compared with the mean	202
Table 46: Theo's listening progress and the cohort mean	232

Table 47: Missing and wrong words: Comparing Theo with the full cohort	243
Table 48: Theo’s translations of J’apprends à jouer and Je reste à la maison.....	246
Table 49: Fran’s listening scores relative to the mean	264
Table 50: Fran’s pseudocognate errors.....	272
Table 51: Missing and wrong words: Comparing Fran with the full cohort.....	276
Table 52: Ryan’s listening proficiency, compared with the mean.....	294
Table 53: Ryan’s discussion of approaches to listening over time.....	312
Table 54: Implications for future research	360
Table 55: Rasch analysis: Screenshots of data set-up to export into RUMM.....	428
Table 56: Engagement scores of all listening tests.....	451

Figures

Figure 1: Field’s model of listening (2008b, 2013:97; 2013:101)	25
Figure 2: The Fuzzy Lexical Representation Hypothesis (Gor et al, 2021).....	27
Figure 3: Learning to listen and listening to learn	30
Figure 4: The proposed complex dynamic system of the present study	52
Figure 5: Second language development as plant growth.....	55
Figure 6: Pilot paused translation showing 'notes' box	78
Figure 7: Paused translation data collection procedure.....	80
Figure 8: Establishing the schema in the listening passage	97
Figure 9: AuralLex introductory screen	98
Figure 10: Part of main AuralLex test screen.....	99
Figure 11: AuraGram preamble.....	100
Figure 12: AuraGram main test.....	100
Figure 13: Preamble to the Metacognitive Listening Awareness Questionnaire.....	101
Figure 14: Sample of main part of Metacognitive Listening Awareness Questionnaire.	101
Figure 15: Data analysis process for Research Question 1	103
Figure 16: Data analysis procedure for Research Question 2	105
Figure 17: Elements and their categories within this chapter	119
Figure 18: Average cohort listening scores over the data collection period.....	123
Figure 19: Percentage of attempts at each time point	123
Figure 20: Track 5 of test D25, demonstrating differing dispositions of participants	124
Figure 21: Strategic risk-taking over time	125
Figure 22: Changes in ‘missed the end’ with low numbers of references.....	129
Figure 23: A slight increase in ‘got the start, missed the end’ code	129
Figure 24: Falling codes for ‘got start missed end’	130
Figure 25: Translations of the sound /abl/ when given the word portable	137
Figure 26: Attempts at translating formidable in tests E3 and D3	138
Figure 27: Translations of adorable in test E3	139
Figure 28: Missegmentation of amuser.....	140
Figure 29: Dealing with liaison in c’est utile	142
Figure 30: Mistranslations of campagne over time.....	148
Figure 31: Translations of ‘vélo’ test D8.....	149
Figure 32: Contextual interference for translation of vélo at T1 and T4	150
Figure 33: Correct translation instantly followed by incorrect translation	154
Figure 34: Number of ‘content word wrong’ errors recorded at each time point	156
Figure 35: Change over time in translating je reste as ‘I rest’	159
Figure 36: Change over time translating cadette as ‘cadet’	160
Figure 37: Phonological and textual errors interact with false friends	160
Figure 38: The appearance of an incorrect I like at each time point.....	163
Figure 39: Default to ‘I like’ cross-referenced with engagement.....	164
Figure 40: Change in function word appearance in test E3	168
Figure 41: Grammatical errors of plurality in test D8.....	170
Figure 42: Translations of j’aime regarder in test D25	172
Figure 43: Reminder of the structure of the system and its subsystems	174
Figure 44: Mean paused translation scores plus trendline	177
Figure 45: Moving min-max paused translation scores: mean of cohort of 80	178
Figure 46: Whole cohort AuralLex scores, with polynomial trendline.....	182

Figure 47: AuralLex moving min-max for the whole cohort.....	183
Figure 48: Mean AuraGram scores with polynomial trendline.....	184
Figure 49: AuraGram moving min-max graph for full cohort.....	185
Figure 50: Listening proficiency, vocabulary and grammar on the same scale.....	188
Figure 51: Whole cohort moving correlation between listening vocab grammar.....	189
Figure 52: Moving correlations between listening vocab grammar.....	190
Figure 53: Means plus trendlines for the four MALQ elements.....	193
Figure 54: MALQ elements with moving min-max.....	193
Figure 55: Basic MALQ relationships against each other (whole cohort).....	194
Figure 56: Relationships between listening and the metacognitive subsystems.....	196
Figure 57: Helen’s listening score compared with the mean, including trendlines.....	202
Figure 58: Helen’s moving min-max listening score.....	202
Figure 59: Helen’s listening proficiency changes according to change point analysis...	203
Figure 60: Helen’s AuralLex scores (compared with mean).....	204
Figure 61: AuralLex moving min-max for Helen.....	204
Figure 62: Helen’s AuraGram scores with reference to the mean, plus trendlines.....	205
Figure 63: Helen’s AuraGram moving min-max.....	206
Figure 64: Vocabulary and grammar charted together.....	207
Figure 65: Helen’s moving correlations of listening with vocabulary and grammar.....	208
Figure 66: Grammatical errors of plurality in test D8.....	214
Figure 67: Helen’s metacognitive scores relative to the mean.....	217
Figure 68: Moving min-max for Helen’s metacognitive scores.....	218
Figure 69: Development of Helen’s metacognitive skills.....	219
Figure 70: Moving correlations of metacognitive subsystems viewed together.....	219
Figure 71: Theo’s listening progress, with trendline.....	232
Figure 72: Moving min-max for Theo’s listening proficiency.....	233
Figure 73: Change point plot for Theo.....	234
Figure 74: Theo’s AuralLex scores compared with the mean.....	235
Figure 75: Theo’s moving min-max vocabulary size.....	236
Figure 76: Theo’s grammar scores plus cohort mean.....	237
Figure 77: Theo’s moving min-max grammar scores.....	237
Figure 78: Theo’s vocabulary, grammar and listening scores on the same graph.....	238
Figure 79: Theo’s trendlines for vocabulary and grammar knowledge.....	239
Figure 80: Moving correlations of vocabulary and grammar for Theo.....	239
Figure 81: Theo’s metacognitive scores relative to the mean.....	248
Figure 82: Theo’s metacognitive scores through a moving min-max lens.....	249
Figure 83: Theo’s metacognitive trajectory.....	250
Figure 84: Theo’s moving correlations between listening and metacognitive systems.....	251
Figure 85: Theo’s moving correlations between listening, translation, problem-solving.....	252
Figure 86: Fran’s listening scores relative to the mean.....	264
Figure 87: Fran’s moving min-max listening scores.....	265
Figure 88: Fran’s change points.....	265
Figure 89: Fran’s AuralLex scores, compared with the mean.....	266
Figure 90: Fran’s AuralLex moving min-max score.....	267
Figure 91: Fran’s grammar scores plus cohort mean, and trendline.....	268
Figure 92: Fran’s grammar scores moving min-max.....	268
Figure 93: Fran’s listening, vocabulary and grammar scores plotted together.....	269
Figure 94: Moving correlations of linguistic skills for Fran.....	269
Figure 95: Fran’s individual metacognitive scores relative to the mean.....	279
Figure 96: Fran’s metacognitive moving min-max scores.....	279

Figure 97: Fran’s metacognitive scores on a single plot: raw scores and trendlines	281
Figure 98: Moving correlations btwn listening and the four metacognitive subsystems	282
Figure 99: Fran’s affective approaches over time	283
Figure 100: Ryan’s listening proficiency, compared to the mean, with trendlines	294
Figure 101: Ryan’s listening proficiency moving min-max	294
Figure 102: Ryan’s change point analysis: no change	295
Figure 103: Ryan’s vocabulary size compared with the mean.....	296
Figure 104: Ryan’s AuralLex moving min-max	296
Figure 105: Ryan’s AuraGram scores compared with the mean.....	298
Figure 106: Ryan’s AuraGram moving min-max.....	298
Figure 107: Ryan’s listening, vocabulary, and grammar on a single chart:	299
Figure 108: Ryan’s Moving correlations between listening and vocabulary / grammar	300
Figure 109: Ryan’s metacognitive scores plus cohort mean, with trendlines	306
Figure 110: Ryan’s moving min-max for the metacognitive subsystems	306
Figure 111: All Ryan’s metacognitive subsystems plotted together:	307
Figure 112: Ryan’s moving correlations for the metacognitive subsystems.....	308
Figure 113: Ryan’s Metacognitive subsystems moving correlations.....	309
Figure 114: Moving correlations between linguistic proficiency vocabulary grammar .	322
Figure 115: All four metacognitive subsystems and their correlations with listening	335
Figure 116: Recap of moving correlations for whole-cohort MALQ	335
Figure 117: Moving correlations between the linguistic subsystems at group level)	339
Figure 118: Moving correlations between listening and vocabulary and grammar:	339
Figure 119: Whole cohort moving correlations for the metacognitive subsystems	340
Figure 120: Case studies moving correlations: metacognitive subsystems	340
Figure 121: Field’s model of listening (2008b, 2013:97; 2013:101)	354
Figure 122: Adaptation of Field’s model of listening according to my participants	357
Figure 123: The Comprehension Approach questions for C11, with answers in red.....	373
Figure 124: Rasch model outputs examples.....	429
Figure 125: examples of 'not attempted'	451

1. Introduction

1.1. Research context

Of the four skills into which second language proficiency is often divided – speaking, writing, reading and listening, listening is the most under-researched. This is hardly surprising, as a receptive skill, like reading, it takes place in the mind and has no obvious natural output to explore. Unlike reading, though, it is ephemeral and time-limited, meaning that it is even harder to investigate. Nearly thirty years ago, Vandergrift (1997) called it the ‘Cinderella’ skill – overlooked and undervalued – and while one might argue that this has become a cliché to wheel out at this juncture, it is still the case that we know far less about listening than we do about the other three skills, simply by dint of its slippery, difficult-to-pin-down nature.

What we do know, too, is based on research largely undertaken at university level, and largely with participants learning English as a foreign language (Plonsky, 2023). While these studies have enormous value and have taught the academy much about the nature of second language listening, it is difficult to know the extent to which the findings made in this context can be carried over to learning environments. Equally, most research into second language listening is cross-sectional, and little is known about how second language listening develops.

1.2. Research objectives, questions, scope

The present study, then, attempts to address questions above by exploring the developmental trajectory of second language listening among beginner learners of French in an English secondary school. Recent work has shown that English students find the

learning of French difficult (e.g. Deckner, 2014; Graham & Macaro, 2008; Milton, 2008) and that listening is the hardest skill (Graham & Macaro, 2008). The numbers of students choosing to study French beyond the compulsory period continues to fall year on year (British Council, 2023; Ofqual, 2023).

Therefore I have two key research objectives for the present study. The first is to add to the body of knowledge about the experience of second language listening for the sake of knowledge itself – about the human brain, about listening, about French listening, about beginner listening. But the more practical objective is that by shedding light on the experience of beginner listening of French within the English context, with all the challenges this brings, we might begin to adapt pedagogy, and resources, to make these parts of the lesson more accessible and by extension to boost learning, engagement and motivation in the subject.

It is to these ends I ask:

1. What do the listening errors of school-aged students of French reveal about a) their listening processes, and b) the development of their listening proficiency over a three year period?
2. How do the relationship between listening proficiency and the linguistic subsystems of vocabulary and grammar, and the metacognitive subsystems of Person Knowledge, Problem-Solving, Directed Attention and Mental Translation, develop over three academic years?
3. To what extent do the answers to the above questions vary between individuals and how do individual patterns compare to those seen at the group level?

By taking a Complex Dynamic Systems Theory approach I am committed to the concept that my research generalises not to population, but to theory. In other words, great care

must be taken not to infer what I have found in my own research to other UK secondary schools. Nonetheless, its aim is to provide food for thought both to the researcher but also to those working in similar settings, and it aims to ask as many questions as it answers.

1.3. Outline of the study

This thesis follows a classic structure. I begin by reviewing the literature in chapter two, which leads me to my research questions. I present three theoretical frameworks which work together to inform the study design: Field's model of listening (Field, 2008b, 2013), the Fuzzy Lexical Representation Hypothesis (Gor, Cook, Bordag, Chrabaszcz, & Opitz, 2021), and Complex Dynamic Systems Theory (Larsen-Freeman & Cameron, 2008).

There follows a discussion of various subsystems which contribute to listening proficiency: I have chosen to contextualise listening through two models of linguistic competence and metacognitive awareness (Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006), and I demonstrate the relationships between these variables and listening proficiency. Measurement of listening proficiency and the inherent difficulties thereof are also discussed, which then leads me to addressing longitudinal research into listening.

At this point the two key elements are brought together: Complex Dynamic Systems and longitudinal research into listening. I grapple with the challenges suggested by the current literature, and propose my research questions.

In chapter three, which sets out my methodology, I begin with a discussion of my positionality before moving onto an overview of my research design and the challenges which might arise as a result. I explain the creation and piloting of my various instruments – a measure of listening proficiency, vocabulary and grammatical knowledge – and how I collected and analysed my data. The chapter ends with a discussion of ethics

and the problems that arose during the study, including the covid-19 pandemic which began half way through my data collection.

Findings are reported in chapters four to nine. Chapter four addresses the first research question, which asked what the errors made by my participants tell us about their listening processes and their development. I show that in the main, the phonology of the input drives an attempt at understanding – in other words, that their lexical representation is indeed fuzzy (Gor et al., 2021), leading to a range of mistranslations. I also demonstrate that grammatical knowledge or parsing was largely bypassed both at word level, where function words tended to be overlooked, and at morphemic level. I conclude that each participant’s translation represents a complex dynamic system at a moment-to-moment level.

Chapter five explores a more ‘classic complex dynamic system over the three academic years of the data collection. It shows the ebb and flow of the relationships between vocabulary and grammatical knowledge and listening proficiency, although one in which vocabulary knowledge is consistently more important than grammatical knowledge. A similar model for four key elements of metacognition – Person Knowledge, Problem-Solving, Mental Translation and Directed Attention – are explored, and show that all are inclined towards moderate correlations with listening proficiency.

Chapters six to nine contain the four case studies: Helen, Theo, Fran and Ryan. They set out to answer the same questions that were addressed in chapters four and five but for each individual only, and demonstrate the wide variation in the experience of learning to listen, with each individual forming their own complex dynamic system as we get to know them. It demonstrates the range of trajectories of learning and how unique each learner’s journey is.

Discussion and analysis is in chapter 10. I split this chapter into two parts: the first addresses the group level findings, and the second the case studies, and I attempt to contextualise my findings within the existing literature, finding some elements which are consistent with current scholarship, but others which challenge it. In this section I have also addressed the impact of the covid-19 pandemic on language learning.

Implications of my findings are addressed in chapter 11 and are divided into theoretical, scholarly, methodological and pedagogical. The thesis ends with limitations, a conclusion, references and appendices.

1.4. A personal note

When I began my MSc at Oxford in 2015, I never thought my journey would take me to a DPhil, and never imagined that if it did, it would be in the developmental trajectory of listening. Still, the presentation of a single study – Graham & Macaro (2008) – peaked my interest as I saw the relevance of the research questions to my own context as a teacher of French in an English secondary school. A Master's dissertation about the experiences of listening to French as a lower intermediate English learner (Simpson, 2017) showed me the excitement of trying to grasp what happens in the listener's mind when they hear French, with all the intriguing nuances that this brings.

A rainy evening in the university library at the start of my DPhil. The full-time students have long since gone home, and another paper captures my imagination. For practical reasons I am doing the DPhil part-time, and continuing to work, and want to make a virtue of my part-time status by looking at the progress of listening over as long as possible. While reading around longitudinal research, I come across Larsen-Freeman (2006) which simultaneously introduces me to Complex Dynamic Systems Theory and brings alive four case studies, telling a story of their learning which inspires me on an

academic and a professional level. As a teacher I had often been left feeling frustrated with research which presented two variables: one dependent and one independent, as I knew instinctively that learning was not that simple. Complexity made much more intuitive sense as I knew the myriad ways in which my own students' learning was impacted. Putting these two ideas together, the present study was born.

Working on this thesis has been a labour of love over nearly seven years. It has been exciting, frightening, surprising, motivating, upsetting, challenging but never boring. It astonishes me that the project is at an end. I hope you find it as intriguing as I have done.

2. Literature Review

2.1. Introduction

This chapter will present the academic literature which is pertinent to the present study. I will address the theoretical frameworks upon which the study is based, and the key components which contribute to successful second language listening. I will also address Complex Dynamic Systems Theory both more broadly and also with respect to listening research. Finally I will discuss research specifically on the subject of error analysis in listening as well as longitudinal research into listening.

2.2. Theoretical frameworks

2.2.1. Theoretical framework for the process of listening

There are several key models of L2 listening: Anderson's (1990) three-phase model is made up of 'perceptual processing' (i.e. hearing a sound as a familiar language), 'parsing' (in which the listener breaks the sounds into words and chunks), and 'utilisation' (i.e. giving meaning to words and chunks). Cutler and Clifton (1999) divide same process into four operations: 'decoding', 'segmenting', 'recognising' and 'utilising' where Anderson's 'parsing' becomes the two elements 'segmenting' and 'recognising'. Field's model (2008b, 2013), illustrated in Figure 1, provides even more detail.

Despite its theorisation as being comprised of phases, the process of listening is not linear, but iterative (Grosjean, 1986; Graham & Macaro, 2008; Vandergrift, 2015). That is to say, as the listener continues to receive new input, hypotheses of what has been understood are being constantly revised, meaning is adjusted, and new anticipations are

made as to what might come next. At any stage in the process, it is possible that breakdowns in understanding will occur (Field, 2003).

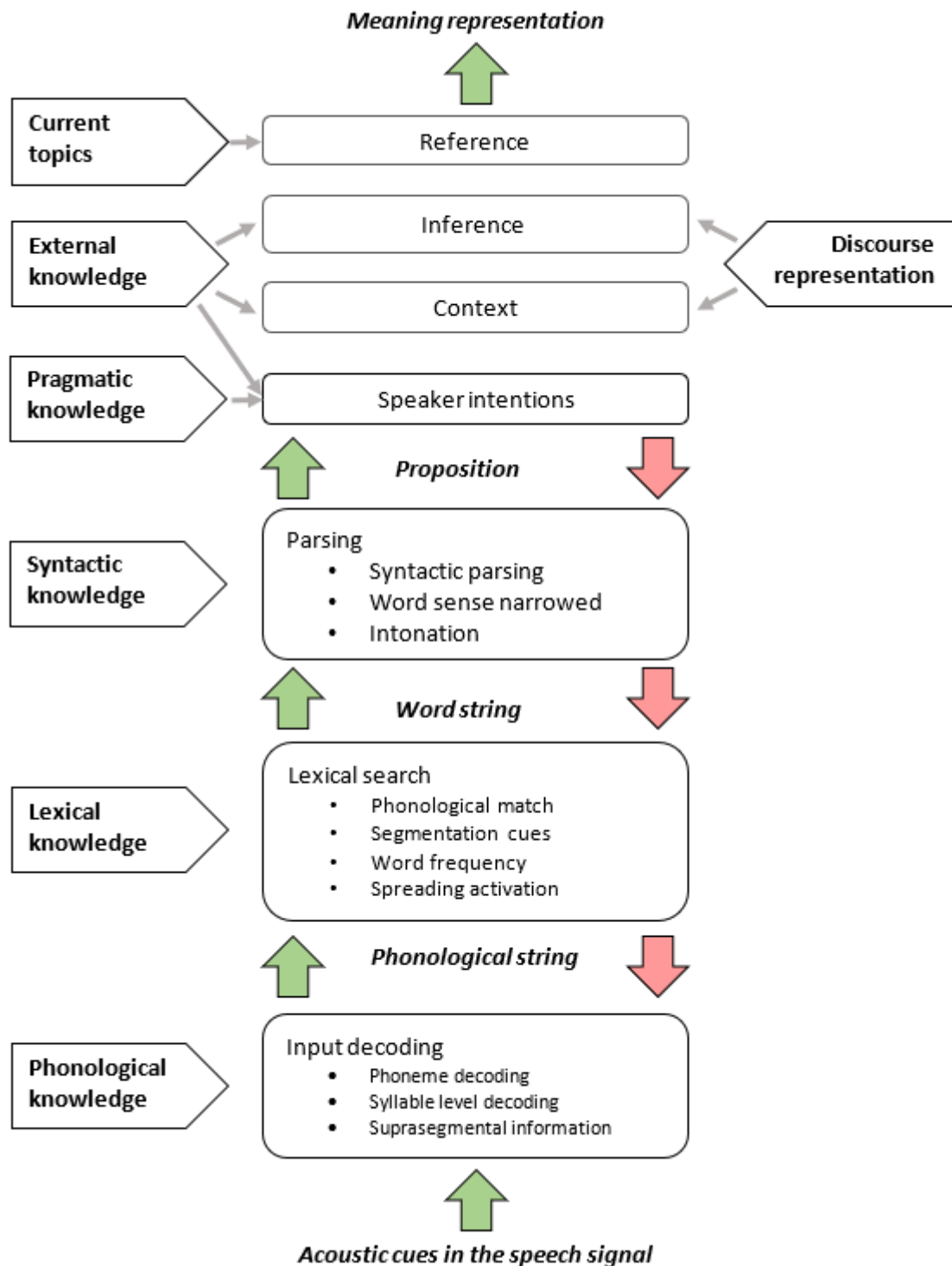


Figure 1: Field's model of listening (2008b, 2013:97; 2013:101)

Calling the processes involved in listening ‘stages’ implies a logical and linear progression from one to the next, suggests Field (2013:94); he prefers ‘levels of analysis, which better reflects the recursive nature of listening’. The present study will be based on Field’s (2013) framework, which is illustrated in Figure 1. It allows me to isolate the different elements in listening in order to examine the relationship between them at different stages in the developmental trajectory.

One key difference for a listener experiencing this process in a second language is the level of automatisisation. While a fluent or native user of a language might largely experience this process without hiccup, a beginner or intermediate L2 listener lacks both the skills and the language knowledge to enable the process to proceed automatically. As a result, the listening process is often compromised, sometimes at several levels of analysis simultaneously.

2.2.2. The Fuzzy Lexical Representation hypothesis

The Fuzzy Lexical Representation hypothesis (Gor et al., 2021) is rooted in the phonological encoding of the second language learner’s lexicon. It builds on previous research such as that of Cook, Pandža, Lancaster, & Gor (2016) who found that second language learners struggle with accurate word storage, resulting in confusion with input as they retrieve incorrect vocabulary; such lexical confusion has also been found by other researchers (e.g. Cook et al., 2016). It speaks to a bottom-up approach on Field’s model: the phonological knowledge (at the bottom of Figure 1) risks being inadequate, the upshot of which is difficulty with identifying words and a particular issue when words are phonologically similar. (Gor et al. (2021) give the example of Japanese learners struggling with ‘rock’ versus ‘lock’.) The hypothesis is set out diagrammatically in Figure 2.

The authors of the hypothesis state that the very existence of the first language interferes with the processing of sounds in the second language, particularly of sounds which are similar to those of the listener's L1. It is this that causes the fuzziness, and by extension, an unreliable form-to-meaning mapping within the listener's mind. This unreliability results in a range of difficulties which are set out in the bottom box of Figure 2, which can be summarised by slow or inaccurate recall of words, driven by their sound more than their meaning.

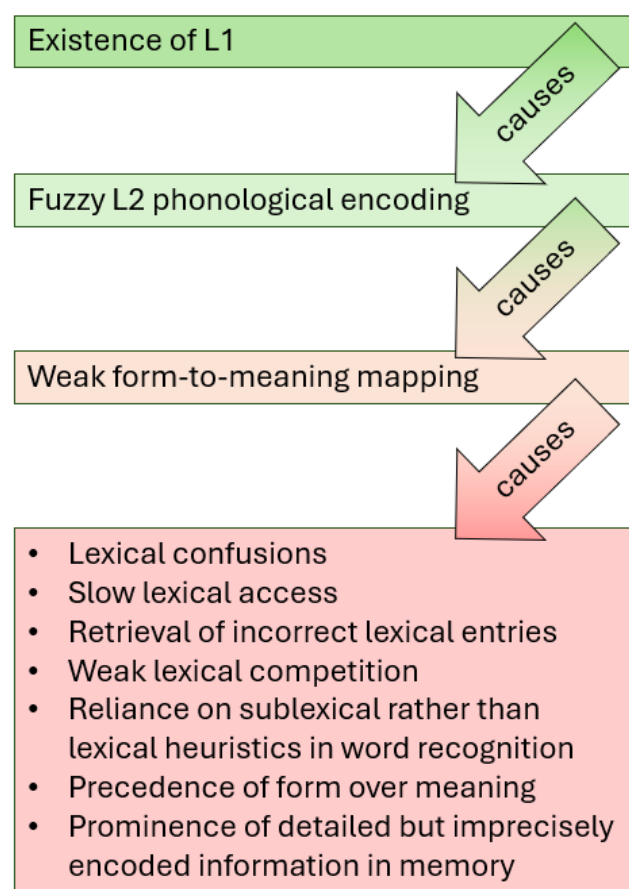


Figure 2: The Fuzzy Lexical Representation Hypothesis (Gor et al, 2021)

While largely phonological, Fuzzy Lexical Representation can also extend to orthography. The inherent lack of clarity in the learner's representation of a word might interact with a variety of other sources of information about the word, such as frequency

or context, and as such, we can see that if the hypothesis were proven, it would impact all stages of Field's model of listening.

The hypothesis has drawn some criticism, notably from Nicol (2022), who disputes Gor et al.'s (2021) assertion that *all* phonological encoding at *all* levels of L2 proficiency is by definition influenced by the L1 and therefore contains levels of fuzziness. Nicol also calls for testing of the hypothesis through spoken production experiments. Nonetheless, she accepts the hypothesis is relevant to beginner and intermediate learners.

Still, the authors and other researchers continue to test the hypothesis, which is gaining momentum. Zhao & Li (2021) for example, posit that the earlier the L2 is acquired, the less fuzzy the L2 representations might be, and by contrast, delayed L2 learning appears to cause increased fuzziness due to the entrenchment of the L1. They also show that this fuzziness can lead to processing errors in the aural modality. Zheng & Gor, (2024) have demonstrated the cross-linguistic influence between an L1 and an L2 and that while difficulties caused by fuzziness were somewhat overcome through directed training (with higher-intermediate learners), the lexical challenges were not completely mastered.

The Fuzzy Lexical Representation Hypothesis is new, but evidence so far appears convincing and highly relevant to the present study.

2.2.3. Listening to learn, or learning to listen?

The model set out in Figure 1 and the accompanying argument in section 2.2.1 concerns itself with the process of listening, and by extension, the context of how a student of a second language develops their listening skills or 'learns to listen'. Although it is largely beyond the scope of the present study, it is important to consider the ultimate purposes of listening to a foreign language. These can be summarised as 'listening to learn'. This

might be listening in order to learn what the speaker is communicating and is as true of listening to a first language as listening to a second (Vandergrift, 2004).

One might similarly contrast ‘listening for comprehension’ and ‘listening for acquisition’ (Richards, 2005): the latter being purpose of listening in order to enhance one’s L2, with (Field, 2008:334) stating that:

‘Listening is the principal means by which learners expand their knowledge of the spoken forms of the target language. It opens up access to language used in natural contexts. It also is a much more effective and efficient channel for picking up grammar and vocabulary.’

To test whether listening really is an ‘effective and efficient channel for picking up grammar and vocabulary’ is to test several key theories about second language acquisition.

Krashen’s (1985) Input Hypothesis asserts that second languages are acquired ‘by understanding messages or receiving comprehensible input’, but Schmidt (1990, 2001, 2010) develops this thinking by suggesting that input would only become ‘intake’ if the learner noticed or attended to the input they were receiving, and that classroom pedagogy, including the demands of a specific task, would direct or influence the noticing.

In 1990, then, Schmidt called for more research into ‘what learners are conscious of’. In other words, to facilitate language learning, the researcher or teacher must be aware of the listener’s experience – what they are conscious of – in order to exploit it into something explicitly teachable or learnable. But at the outset, the beginner or intermediate listener (illustrated in red below) might not be able to ‘listen to learn’, because the input is not accessible enough. It is important for both teachers and

researchers to know ‘what learners are conscious of’ in order to reach the sections of the diagram illustrated in orange – in other words, to create achievable tasks, or to facilitate appropriate motivation or proficiency. It is only at this juncture that the listener is in a position to whereby ‘listening to learn’ can be activated. The relationship between the two concepts ‘listening to learn’ and ‘learning to listen’ is illustrated in **Error! Reference s** **ource not found..**

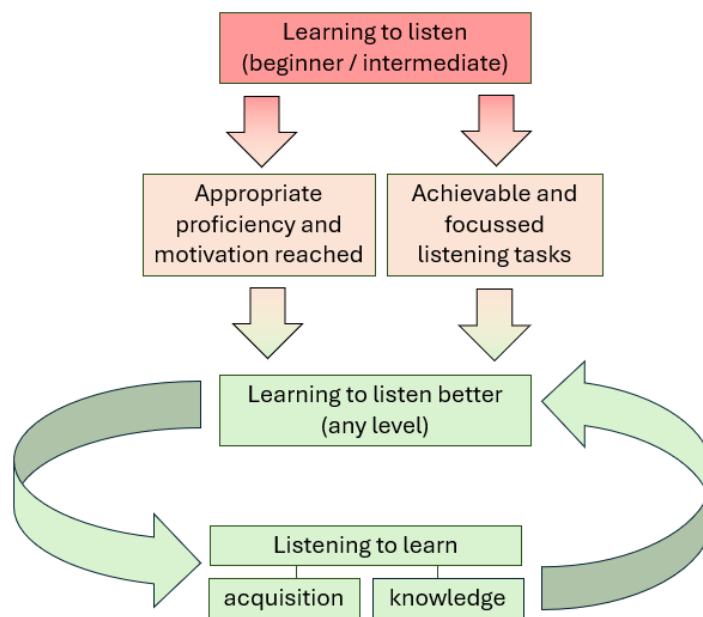


Figure 3: Learning to listen and listening to learn

The present study concerns itself with *how* the L2 learner listens, but this knowledge only has application if one underpins it with *why* the L2 learner listens. It is important that both teachers and researchers do not take for granted the initial stage of the diagram in Figure 3. In other words, for listening in the language classroom (or other L2 learning setting) to achieve its aims, those with interest in those aims (be they teachers or researchers) need to be aware of the process of listening.

2.3. Measuring listening proficiency

In order to know *how* the L2 listener listens, it is helpful to learn *how much* the listener understands. Yet as the spoken word is fleeting, and meaning-making takes place in the mind, this is not a straightforward task. In this section I will set out the different approaches taken by both teachers and researchers to measure listening proficiency, and some of the advantages and pitfalls accompanying them.

2.3.1. The Comprehension Approach: the product of L2 listening

The comprehension approach (Field, 2008b) is when a passage is played to a listener, and followed up with questions. The way in which the listener answers the questions is then taken to be indicative of the listener's level of understanding. A score or mark is frequently given.

With listening comprehension a popular way for both teachers and researchers to ascertain the extent to which a learner has understood a text, there is an emphasis on the 'product' of listening over the 'process': that is, the answers to the comprehension test become the operationalisation of 'listening performance'. Indeed, given that listening takes place in the mind, it might be argued that this is a reasonable, overt representation of understanding. However, this kind of test leads to inevitable distortions of the process of listening. Firstly, the nature of the questions is likely to guide the listener to focus on particular elements of the passage; questions can be manipulated at a range of proficiency levels which might give a scorer an unrealistic impression of the level of understanding; correct answers given by the listener might be the result of guesswork or general knowledge as much as a response to the input, and comprehension questions rely also on reading skills, working memory, and are not representative of the real life experience of L2 listening (Prince, 2014; Field, 2011).

An alternative option is ‘free recall’, in which a listener listens to an audio track, and is asked to then recount (in L1 or L2) as much as they can remember of its contents. Here, however, the confounding variable is the listener’s writing or speaking ability, and therefore free recall presents a different type of distortion (Prince, 2014), and working memory capacity of the listener still needs to be considered (Leclercq & Edmonds, 2014). This said, for many purposes, being able to recall the content of what was said afterwards is part and parcel of ‘real life’ listening proficiency.

Thus we can see that both ways of measuring how much has been understood – comprehension questions or free recall – both require a certain level of literacy of the listener (unless relayed orally), and results inevitably measure other elements of cognition and L2 knowledge (Tracy-Ventura, McManus, Norris, & Ortega, 2014).

A third option is that of dictation, which tests a listener’s ability to decode and segment stretches of speech. Asking a listener to transcribe only short chunks at a time should remove the working memory confound (Yeldham, 2017b; Field, 2019). In the ‘paused transcription’ exercise (Field, 2008b; Yeldham, 2017b), an audio track is paused, and the listener is required to write what they have heard. This could be what they have heard in its entirety, or a certain number of words (Estes, 2014).

Taking the dictation construct further, the listener can write what they have heard either in their L1, or in the target language, or both. While Field (2008b) advocates for same-language transcription, he acknowledges that this then begins to tap into an assessment of L2 writing proficiency. Listeners with good working memory are also at an advantage given that they might be able to write what is retained in their echoic memory, without demonstrating any understanding of what they have heard. Paused translation, then, ensures that a research participant is demonstrating a level of understanding rather than

simply ‘parroting’. One might counter that paused translation introduces interference of the participant’s first language into the data that might be generated. Yet in section 2.2.2 I addressed the Fuzzy Lexical Representation hypothesis, which is founded on the premise of the omnipresence of the L1 in L2 listening; if this premise is accepted, such interference is part of the L2 listening experience and as such is not a confounding variable.

2.3.2. Error analysis in listening

A key approach to understanding the listener’s experience – and therefore by extension their developmental trajectory – is to analyse the errors made when listening. Such an analysis depends first on operationalising how listening is measured or assessed. As I discussed in section 2.3.1, options could be to examine answers from comprehension tests, but this is unlikely to generate much meaningful information without accompanying stimulated recall (whereby the listener is asked to explain their thinking). Another option is to ask the listener to do a paused translation or paused transcription task, which will offer insight into their thinking processes.

Much research, then, takes advantage of paused translation or paused transcription to analyse errors given that they can provide insight into the kinds of challenges faced by listeners (Rost, 2016). Wong, Leung, Tsui, Dealey, and Cheung (2021) focussed on the phonological errors produced by Cantonese undergraduates learning L2 English. In their study, data was collected and subsequently categorised qualitatively into five groups (consonant errors / vowel errors / consonant AND vowel errors, changes in syllable numbers, metathesis). Although the focus was bottom-up difficulties, they recognised that both affect and anxiety might have contributed to the data generated.

Sheppard and Butler (2017) took a qualitative and recursive approach to the analysis of their participants' decoding errors (L2 English, with L1s of Chinese, Japanese or Arabic), aiming to explore the influence of text difficulty, content versus function words and articulation rate. They identified that participants struggled to transcribe texts which had been expected to be accessible, and that participants favoured content words over function words, but that articulation rate did not appear to influence quality of decoding. Given that their participants had been resident in the USA for 11 months at the time of data collection, the difficulty they continued to have with listening is noteworthy.

By contrast with Sheppard and Butler's (2017) longer texts, the Japanese EFL learners in Lange and Matthews' (2023) study were exposed to three-word chunks only. They were asked to complete a paused transcription task followed by stimulated recall. Post-hoc qualitative analysis of the paused transcription data gave rise to four identified trends: limited collocation familiarity, syntactic knowledge constraints, difficulties using co-text, L1 phonological influence. Unlike the previous studies I have cited, that of Lange and Matthews is the first to identify some top-down difficulties through a paused transcription protocol. The researchers state that their participants struggled with stimulated recall and demonstrated minimal introspection skills: they suggest that this might be because the listening process more broadly does not lend itself well to analysis; however, the shortness of the extracts might well have exacerbated this situation.

Three papers (Carney, 2021; Cross, 2009; K. R. Leonard, 2019) analyse listening errors through use of both a recall protocol (that is, listening in the L2 then 'free writing' in the L1) and a paused transcription. In all three, participants are advanced L2 learners, and the focus is decoding errors. I cited in section 2.2.2 a classic decoding error with L1 influence where a Japanese learner confuses the English words 'lock' and 'rock'. Cross (2009) cites a similar error where the word 'carrot' in the English phrase 'a carrot to

promote recycling' presents almost universal difficulties among his participants, who almost all attempt to translate it as 'character'. While it is reasonable to categorise both as 'decoding' errors, without stimulated recall and careful probing of each individual who had made the error, one cannot be absolutely sure that phonology or bottom-up processing was definitely the cause of the error. Returning to the theoretical frameworks addressed in section 2.2, it could have arisen because of fuzzy lexical representations of the two words concerned. But through the lens of Field's model of listening in Figure 1 (Field, 2008b, 2013), it might also be reasonable to suggest that both these cited errors stem from a range of the top-down stages of the model, including inference, context or speaker intentions. Care must be taken, then, not to place errors simply in one category without acknowledging the sometimes complex interplay between many of the stages of analysis.

Furthermore, thinking methodologically, the very nature of the introduction of pauses in a paused transcription test might interact with the listening process. This might create a bias towards bottom-up processing: by not being exposed to the full text, the listener is less able to form an overview of its content and therefore is less able to deploy some of the top-down approaches to making meaning.

2.4. Listening and subsystems

We can see from Figure 1 that a range of variables, or subsystems, work together to enable listening to take place. While there is a considerable body of work which concerns itself with the relationship between L2 listening proficiency and single subsystems, with the exception of the studies by Vandergrift and Baker (2015; 2018), none have attempted to consider how the subsystems interact with one another: given that this interaction is the key to listening success (Buck, 2001), this appears to be a clear gap in the

literature. This section will provide an overview of the current understanding of the relationships between L2 listening proficiency and the key subsystems listed in Figure 1.

2.4.1. Listening and vocabulary knowledge

There is a strong body of evidence correlating vocabulary knowledge with L2 listening, although the relationship between the two is not entirely straightforward. To start with, in order to access the vocabulary they know, listeners must be able to identify the words within the stream of speech, which they sometimes find challenging (Field, 2003): thus auditory discrimination will confound the findings: for example, when comparing the role of vocabulary knowledge in reading and listening proficiency, Mecarty (2000) found that vocabulary knowledge (operationalised by written word-association and word-antonym tasks) predicted about 25 per cent of reading ability but only about 14 per cent of listening ability, with a suggestion that the demands of working memory and segmentation skills in listening account for the difference between the two.

Nonetheless, vocabulary knowledge has emerged as a key predictor for L2 listening success both among very proficient listeners (Stæhr, 2009) and beginners (Vandergrift & Baker, 2015; Matthews, 2018). However, when considering vocabulary knowledge in L2 listening through the lens of individual differences, van Zeeland and Schmidt (2013) found that, when contrasted with L1, the contribution of vocabulary for L2 was more varied, which suggests that some L2 listeners cope better with unknown vocabulary than others: this might be related to strategy use to compensate for lack of linguistic knowledge (Stæhr, 2009; Bonk, 2000; van Zeeland, 2014).

The field of vocabulary knowledge can also be broken down according to word class, and there is a body of research which addresses the relative difficulty of content and function words when listening. For example, Field (2008) demonstrated that intermediate listeners

of English preferred content words to function words, and Graham and Santos (2013) found that English learners of French struggled to discern function words. Similarly Sheppard and Butler (2017), participants, who were learning English with L1s of Chinese, Japanese or Arabic, found function words more challenging. Taking a paused transcription approach, their participants correctly transcribed 76% of content words but only 54% of function words. Findings were consistent in a similar study, whose participants were also intermediate and advanced listeners of L1 Chinese and L2 English (Yeldham, 2016).

It is reasonable to say, then, that the well evidenced relationship between listening proficiency and vocabulary knowledge might well be a starting point from which other relationships stem: in this section we have seen suggestions that from vocabulary knowledge come relationships with listening and segmentation skills, listening proficiency and grammar, listening and strategy use, and listening and working memory.

2.4.2. Listening and segmentation skills

Being able to hear the gaps between words – segmenting the speech stream – is a crucial and very basic element of L2 listening, yet one that presents many problems for learners. Goh (2000) found that this was a key listening problem among 40 Chinese students of English, and intervention studies have shown how teaching of segmentation can improve listening proficiency (Graham & Macaro, 2008; Simpson, 2017). In a similar vein, Vandergrift and Baker (2015) measured auditory discrimination as a factor in successful L2 listening, and found it played a significant role. However, it is difficult to determine the extent to which segmentation plays a factor, as it is inevitably so closely connected to vocabulary knowledge: good segmentation skills allow listeners to activate their lexical knowledge (Goh, 2000; Graham, Santos, & Vanderplank, 2010).

Questions must also be raised as to the grammatical or perceptual root of segmentation problems. Carroll (2004) suggests that problems associated with segmentation are grammatical rather than perceptual, and here contrast must be drawn with the work of Al-Jasser (2008), where the teaching of English phonotactics – a perceptual rather than grammatical approach – was also found to significantly aid segmentation skills.

Kennedy, Guénette, Murphy, & Allard (2015) also infer that the comprehension problems experienced by their participants are perceptual, citing in particular segmentation issues when several multi-syllabic words occur in succession. There might be linguistic reasons for this: French, as a syllable-timed language (Lidji, Palmer, Peretz, & Morningstar, 2011) poses particular segmentation issues for speakers of stress-timed languages, such as English (Cutler, 2002).

Therefore, we can see that successful segmentation is a key element in listening proficiency, but its impact is not clear, and that it works in tandem with other elements, crucially vocabulary, often strategy use and possibly grammar.

2.4.3. Listening and grammatical knowledge

In section 2.4.1 I discussed that the body of literature which addresses the role of content and function words in listening, which appears to suggest that although they are more frequent, function words are also less salient within the speech stream and therefore can be difficult to detect. It is reasonable, then, to hypothesise that grammatical morphemes are going to present a similar challenge in a more extreme form (N. Ellis, 2017).

Despite the key existence of grammar or parsing on Field's (2008b; 2013) and others' models of listening, research is relatively sparse and inconclusive. One notable study is that of Mecartty (2000), which found that the relationship between grammatical knowledge and L2 listening was not significant, although it was for L2 reading. Mecartty

suggests that this might be due to the overlap between grammatical and vocabulary knowledge, as well as the unsatisfactory nature of her Comprehension Approach to L2 listening. She states that the comprehension questions were a combination of ones that focussed on ‘main ideas’ and ‘details’ – however these questions are not given in the appendix and it is possible that, to answer the questions correctly, knowledge of grammatical constructions was not necessary. Andringa, Olsthoorn, Van Beuningen, Schoonen, & Hulstijn (2012), and Vafae & Suzuki (2020) also detected some relationship between the two variables, with their choice of instrument a grammaticality judgement test. A sentence construction task is an alternative instrument, whereby the participant is given a sentence starter and two possible options for an appropriate grammatical continuation. When using this instrument, Hui & Godfroid (2020) found no robust contribution of parsing skills to L2 listening.

In contrast with the evidence on a clear relationship between listening proficiency and vocabulary, then, it is harder to state with any certainty that either good grammatical knowledge contributes to listening proficiency, or vice versa, due to both the paucity of papers on the topic, as well as the inconclusive results of the papers that have been published, when taken together.

2.4.4. Listening and metacognition

There is a wealth of literature on the role of metacognition in listening – that is, listener awareness of the cognitive processes involved in comprehension, and the capacity to oversee, regulate and direct these processes (Vandergrift & Goh, 2012), although the evidence derived from this is not always straightforward.

Although the importance of metacognition (e.g. Cross, 2010, Bozorgian, 2012, C. Goh & Taib, 2006, Vandergrift & Tafaghodtari, 2010), is somewhat of a recurring theme in the

L2 listening research, Graham, Santos and Vanderplank (2010) suggest that care should be taken before classroom teachers conclude that metacognition is the only way to improve L2 listening. Still, gaps in both knowledge and understanding are inevitable for L2 listeners, and as a result, a certain level of strategy use is needed to fill these gaps as well as possible.

The principle that skills are needed before strategies can be deployed (Macaro, 2006) might explain why Vandergrift and Baker (2015), researching 13-year-olds at the end of their first academic year of French immersion in Canada, found that metacognition was a strong predictor in L2 listening success, but the same researchers (Vandergrift & Baker, 2018) with 10-year-olds only two months into their first year of French immersion found the same link to be weak and non-significant. This discrepancy in findings might also suggest that as well as a linguistic proficiency threshold (Schoonen, Hulstijn, & Bossers, 1998; Dong, 2016), use of strategy requires a certain age or maturity, although Goh and Kaur (2017) reported success with metacognition training with younger children.

In their Metacognitive Awareness Listening Questionnaire (MALQ), Vandergrift, Goh, Mareschal, and Tafaghodtari (2006) identify the following elements: Person Knowledge, Problem-Solving, Mental Translation and Directed Attention. The questionnaire, with the wording of the questions for each element, can be found in appendix 15.1.1. In an earlier paper (2003), Vandergrift references the concept of ‘orchestrating’ strategies – in other words, the idea that L2 learners need to make use of strategies in effective combinations for specific tasks. However, Wang and Treffers-Daller (2017) found that among their participants (Chinese university students who were not majoring in English), the only element to have a significant relationship with listening proficiency was Person Knowledge.

Indeed, Person Knowledge might be argued to push the boundaries of our definitions of metacognition. The questions tapping into this part of the MALQ are: ‘I find that listening in [language] is more difficult than reading, speaking, or writing in [language]’, ‘I feel that listening comprehension in [language] is a challenge for me’, and ‘I don’t feel nervous when I listen to [language]’. Such questions appear to address constructs more consistent with those of self-efficacy and anxiety: the broader construct of the ‘affective filter’ which accounts for how variables such as motivation, attitude, anxiety and self-confidence might influence language-learning (Dulay & Burt, 1974).

Literature is largely consistent that listening anxiety is associated with reduced proficiency. For example, Brunfaut and Révész (2015) found that less anxious listeners performed better in their study of non L1 English undergraduates at an English university. Vafae and Suzuki (2020), with participants at Iranian EFL classes, showed that anxiety was a significant predictor of listening ability. Xu (2011) makes a connection between task type and levels of anxiety, suggesting that the Comprehension Approach increases anxiety; this finding is reinforced by X. Zhang (2013), who also found that listening anxiety was situation-specific. However, when researching American college students learning French, Mills, Pajares, & Herron, (2007) found that self-efficacy was a stronger predictor of performance than anxiety. Zhang (2019) returned to the topic with a meta-analysis, finding through 55 different papers a strong correlation between listening anxiety and performance, across a range of proficiency level, although there was some suggestion that both cross-linguistic similarities and youth might mitigate anxiety.

Self-efficacy – when a person is confident that they can complete a task (Bandura, 1993) – is probed with the questions about difficulty and challenge of L2 listening, and there is

a broad body of literature connecting greater self-efficacy with better listening proficiency (e.g. Graham, 2007; Graham & Macaro, 2008; Lanvers, 2020), although Mills et al. (2007) found only that it predicted self-regulation, not academic achievement. When examined in a broader academic context, a meta-analysis found it to account for 14 per cent of variance in academic performance (Multon, Brown, & Lent, 1991).

Mental Translation

Mental Translation¹ is a strategy that ‘beginning-level listeners often feel compelled to use’ (Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006:450), while acknowledging that it is counterproductive. Still, one cannot escape the fact that any learner of a second language inevitably has a first language and there is a wide body of evidence which suggests that the influence of the L1 is ever-present while processing the L2. (e.g. Hall, 2002; Nation, 2000; Wilcox & Medina, 2013; Broersma & Cutler, 2008).

Nonetheless, although researchers have identified mental translation as unhelpful during listening (Bloomfield et al., 2010), it is unclear whether a reduction in mental translation is simply indicative of growing listening proficiency, or whether other factors are at play.

Directed Attention

In section 2.2.2 I addressed the Noticing Hypothesis (Schmidt, 1990, 2001, 2010), which states that ‘input’ can only become ‘intake’ through conscious attention. The Directed Attention section of the MALQ asks the participant to rate the following statements:

- I focus harder on the text when I have trouble understanding

¹ I capitalise Mental Translation when referring to the specific subsystem measured by the Metacognitive Awareness Listening Questionnaire. When I am referring to mental translation generally, it is not capitalised.

- When my mind wanders, I recover my concentration straight away
- I try to get back on track when I lose concentration
- When I have difficulty understanding what I hear, I give up and stop listening

Rost (2016:171), then, defines a series of different types of attention, which include Directed Attention (‘consciously ignoring distractions or a tendency to give up’), Persistent Attention (‘keeping flow of attention even if temporarily distracted by unknown language’) as well as Noticing Attention (‘attending to new language and specific language’). It follows that Noticing Attention can only take place if the listener succeeds in consciously ignoring distractions, including those caused by unknown language: the listener who has been distracted or who has given up is not able to Notice new language.

Even listeners who are motivated to persevere and resist distractions might struggle with attentional issues. For example, when listeners encounter unknown words in reading, there is evidence that attentional problems are created (Graves, 2009; K. Nation, Snowling, & Clarke, 2007); in listening, neuroscientists have detected a time-lag of 400 to 600 milliseconds between input and processing of unfamiliar input (B. Liu, Wang, & Jin, 2009; Seyednozadi, Pishghadam, & Pishghadam, 2021), which is long enough, in turn, to delay subsequent Noticing opportunities.

Both theory and research, then, indicate that attentional issues are likely to present difficulties for beginner second language listeners, with both linguistic and affective consequences.

Metacognition in literature: summary

To sum up, plenty of evidence confirms the existence of a positive relationship between metacognition and L2 listening success in general, although the evidence is less clear for

younger learners. However, it is important to see metacognitive strategies – as portrayed by the MALQ – not as a discrete set of approaches to second language listening, but as one way to conceive of a much wider range of overlapping factors and constructs, all of which contribute to (or detract from) successful second language listening proficiency.

2.5. Longitudinal research in L2 listening

Longitudinal work into second language acquisition has tended to take one of three designs: (1) descriptive-quantitative studies of L2 development, (2) research on L2 programme outcomes, and (3) investigation into instructional effectiveness (e.g. intervention studies) (Ortega & Iberri-Shea, 2005).

Table 1, below, cites the published research undertaken longitudinally on L2 listening. Through thorough database searches and citation-chasing, efforts were made to include all papers published since 2000, plus Thompson and Rubin's influential 1996 work. As can be seen, the vast proportion of longitudinal work on L2 listening falls into the category of intervention studies (column 'I'). It is also evident from the table that strategy use (column 'S') forms the core of this work.

2.5.1. Strategy use: the process of L2 listening?

Given the preponderance of strategy use research in longitudinal L2 listening research, it is worth considering its place in the process-product debate. 'Product', in the form of some kind of comprehension test, is a common a pretest / posttest instrument. Yet, to elicit participants' strategy use, other instruments, such as questionnaires or other forms of self-report are often used. When examined carefully, these instruments might be able to shed light on the *processes* involved with listening. Careful reading of the strategy use literature appears to suggest that one of its aims is indeed to better understand the process

of L2 listening. For example, both Yeldham (2017) and Dong (2016), imply that analysis of strategy use is a vehicle for understanding progress in listening, and Vandergrift and Goh (2012:99) state that one of the purposes of strategy instruction research is to ‘understand the nature of L2 listening and the demands of learning how to listen’ – as if strategy use is a lens through which we discover how people learn to listen. Graham, Santos and Vanderplank (2011:436), similarly, state, in their paper entitled ‘Exploring the relationship between listening development and strategy use’, that their work has ‘the ultimate aim of improving our understanding of how listening ability in a foreign language develops’.

Yet, examining listening through the lens of a strategy use questionnaire presents as many distortions as does examining listening through comprehension tests (however they are operationalised). For example, strategy use might be constrained by knowledge of available options, or task type (Macaro, 2006): strategy training has been shown to enhance bottom-up skills such as word boundary identification (Graham & Macaro, 2007, 2008), or to boost metacognition and to better access top-down approaches to listen (e.g. Bozorgian, 2012; Cross, 2010; Goh & Taib, 2006; Lotfi, Maftoon, & Birjandi, 2016; Vandergrift & Tafaghodtari, 2010).

In sum, then, the findings from the listening strategy literature are relevant, and much of it sets out to discover whether strategy training can improve listening skills. It is inevitable that L2 listeners will have gaps both in knowledge and understanding of the auditory input, and to fill those gaps the listener will use a range of strategies (Field, 1998, 2011). Graham, Santos, & Vanderplank (2010) acknowledged that listening teaching in England needs improvement, and identify the role of good strategy use as a starting point to enhance linguistic skills.

Table 1. Summary of recent longitudinal research into L2 listening

Author / date	Research Questions	Intervention?	Strategy?	length
Al-jasser, 2008	What is the effect of awareness raising for English phonotactics? Is the resultant knowledge applied to lexical segmentation? How automated does the process become?	y		8 weeks
Becker, 2015	Does the use of explicit metacognitive instruction lead to an improvement in listening comprehension scores? Does metacognitive instruction aid the automatization	y	y	8 weeks
Bozorgian, 2012	Where does metacognitive instruction benefit less-skilled EFL listeners' comprehension?	y	y	4 lessons
Chang & Millett, 2014	Reading only / reading while listening / listening only - which gets best LC results at the end?			13 weeks
Ai-hua Chen, 2009	How do EFL learners adjust their initial learning strategy use over time during a course of listening strategy instruction?	y	y	14 weeks
Chen, 2013	How, if at all, do EFL learners cope with their listening problems over time as they heighten their strategy use?	y	y	14 weeks
Chenjun & Li, 2012	Does explicit instruction and training of certain decoding skills improve L2 listeners' general listening comprehension performance?	y		6 weeks
Cross, 2010	What is the role of peer-peer dialogue in raising learners' metacognitive awareness of L2 listening?	y	y	5 lessons
Dong, 2016	What are the dynamic developmental patterns of the EFL learner's use of listening strategies influenced by listening strategy training? What is the dynamic correlation between the learner's listening strategy use and listening performance over the observation period?	y	y	42 weeks
Goh & Taib, 2006	To elicit and identify the primary school pupils' metacognitive knowledge about listening in English. To investigate how useful process-based activities were for teaching listening.	y	y	8 lessons
Graham, 2007	To investigate how far the self-efficacy of year 12 learners of French could be improved by a programme of strategy training.	y	y	6 months
Graham & Macaro, 2008	Can a programme of strategy instruction improve listening comprehension? Is the level of scaffolding provided a factor in improvement? What are the effects of strategy instruction on self-efficacy beliefs?	y	y	6 months
Graham, Santos, & Vanderplank, 2008	How does listening strategy use develop or change over time in the absence of explicit strategy training? How does strategy use develop over time in students who score differently on a listening test?		y	6 months

Graham, Santos, & Vanderplank, 2011	In the absence of listening strategy instruction, do listeners remain in the same listening proficiency group after six months? To what extent is movement or non-movement between listening proficiency groups after six months related to change in strategy use? To what extent does learners' strategic behaviour reflect their teachers' approach to listening?		y	6 months
Kennedy & Blanchet, 2014	Do FSL learners in Quebec who receive formal instruction about a) familiar and B) standard connected speech processes improve their perception of connected speech? How does the quality of learners' language awareness relate to their development of connected speech perception?	y		15 weeks
Levak, 2013	What are SL learners' perceptions of using Second Life and Skype for developing listening skills in their target language?	y		9 months
Lotfi, Maftoon, & Birjandi, 2016	Can listening strategy training enhance pre-intermediate and intermediate EFL learners' listening performance? Do different approaches to listening strategy training have different effects on performance? How do participants evaluate the impact of strategy training?	y	y	16 weeks
Mareschal, 2007	What are the effects of a self-regulatory approach to L2 listening instruction? Are there interrelationships between language learners' metacognitive awareness, self-regulatory abilities, listening comprehension strategy use and their overall success in listening comprehension?	y	y	12 weeks
Nogueroles Lopez, 2017	How the frequency of typical strategic behaviour develops after the course of the strategies in instruction	y	y	1 semester
Rahimirad & Shams, 2014	What is the effect of teaching metacognitive strategies on the listening performance of Iranian university students? What is the effect of metacognitive strategy instruction on students' metacognitive awareness of listening?	y	y	8 weeks
Thompson & Rubin, 1996	Can systematic instruction in the use of a range of cognitive and metacognitive strategies result in improvement of listening comprehension?	y	y	2 years
Vandergrift & Tafaghodtari, 2010	Will metacognitive instruction improve listening comprehension? Will the less-skilled learners demonstrate the greatest improvement in listening comprehension? Will the less-skilled learners demonstrate the greatest growth in metacognitive awareness of listening?	y	y	1 semester
Yeldham & Gruba, 2016	How do individual learners develop in an L2 listening strategies course?	y	y	21 weeks
Yeldham, 2016	Will listening comprehension develop more in the interactive course or the strategies course? Will listening strategies and affect-related learner characteristics improve more in the strategies course or the interactive course? Will bottom-up skills develop more in the interactive course or in the strategies course?	y	y	22 weeks

2.5.2. Multi-wave longitudinal research: introducing CDST research

Much of the longitudinal work on listening (cited above in Table 1) operationalises development as a Time 1 test, and, ideally, an improved score at Time 2. Yet to map the developmental path of learners, more detail and nuance is needed. To form such a study, the researcher needs a multi-wave data collection protocol of not only listening proficiency and associated strategy use, but also other subsystems which contribute to the process of listening. This is the model associated with Complex Dynamic Systems Theory (CDST).

One example of such research is Dong (2016), who met with a single intermediate Chinese learner of EFL every two weeks over a 42-week period, in order to document development in both listening performance and strategy use. Unfortunately, the very detailed analysis of many different strategies is not matched by a similarly detailed analysis of any of the sub-skills that make up listening performance; instead, Dong simply records scores on a comprehension test, triangulating it with the participant's diary entries, which, again, emphasise strategy use over skills. The data even suggest that at times, the participant's keenness to use strategy might have been counterproductive in her listening performance, as her attention was focussed on the strategy inventory rather than on what she was listening to.

Despite these concerns about Dong's study, it remains an important work with regard to the present study for a number of reasons. Firstly, it provides a model for very regular data collection with regard to listening performance, in contrast to most longitudinal work, which consists of pretest and posttest, and sometimes a delayed posttest (although Vandergrift & Tafaghodtari (2010) is a notable exception here, with a data collection midway through the intervention). Secondly, by virtue of the regularity of data collection,

Dong's study begins to shed light on the fluctuations in performance that are experienced by a single learner, and hints at the idea that progress is not linear, and indeed that there might be reductions in performance just before a definitive leap forward is made. At the outset of my doctoral project, it was the only piece of work in the current body of applied linguistics literature which applies CDST to receptive language skills.

Since then, four other similar works have been published, including one doctoral thesis. In her doctorate, P. Chang (2017) explored the relative contributions to listening development of language learning aptitude, working memory capacity, motivation, metacognitive awareness and learning styles over time. Her participants were L1 Mandarin Chinese and L2 English at a university in China, 300 of whom took part in the quantitative strand of the research, with ten case studies, over one academic year. She showed that working memory and language learning aptitude (operationalised by the LLAMA test, which assesses vocabulary learning, sound recognition, sound-symbol correspondence and grammatical inferencing) were effective in predicting listening success as operationalised by the Chinese College English Test (CET). Unfortunately she does not discuss why she chose aptitude – that is, *potential* in language learning – rather than actual proficiency in the same measures (vocabulary, sound recognition etc).

Furthermore, due to limitations in her structural equation modelling she was unable to show the ebb and flow of individuals' within the quantitative strand of the research.

Nonetheless, the work builds on that of Dong (2016) to demonstrate the interconnectedness of many variables within a larger system of listening proficiency.

Two connected papers are P. Chang & Zhang (2021), which delves into variability in listening development alone, and P. Chang & Zhang (2023), which concerns itself with the variability of language aptitude and working memory in listening development. P. Chang & Zhang (2021) follows three learners over 43 months and explores the factors

which appear to contribute to progress in their listening (attractor states) and those which detract from it (repellor states) Their work indicates the range of contributory factors to listening progress, be they external, such as test scores, school holidays, or internal, such as emotions. In their 2023 paper, the authors show for that their three participants (again over 43 months), large differences occur both at inter- and intra-individual levels, and demonstrate that within the context of L2 listening, aptitude and working memory are dynamic not static personal traits.

On a shorter timescale, Khaing (2023) took a Complexity approach to development of listening strategy but with only six data points over a two week period. It is noteworthy that even over such a short time frame, both strategy use and listening proficiency were found to fluctuate, and like Dong (2016), Khaing's participant was sometimes inclined to overuse strategy at the expense of attention to listening itself. It goes some way to demonstrate longitudinal listening work over a different time scale with multiple data points.

Ducker (2022, 2024) demonstrates how multi-wave CDST research into listening can work on an even smaller timescale in his projects researching moment-to-moment complexity when students listen to a single hour-long lecture in L2 English. Participants were given a foot pedal in which they were able to log any moment during which their comprehension was compromised, and a post-hoc stimulated recall protocol enabled the researcher to analyse the myriad challenges faced by the listeners at time points dictated by them. Both papers illustrate the range of challenges faced by listeners from both a bottom-up and a top-down perspective and illustrates a volatility of comprehension on a very different time scale to those illustrated on the research by P. Chang, (2017), Dong (2016) or Khaing (2023).

These papers demonstrate a new approach to multi-wave or longitudinal work: one in which multiple data points illustrate variability and conceive of a range of different contributory factors to proficiency in listening.

2.6. Longitudinal research within a CDST framework

At this juncture it is important to consider in more detail how using a Complex Dynamic Systems framework enhances the longitudinal study of L2 listening development, as well as the potential problems which might result.

Complexity Theory has found its way into applied linguistics via mathematics, the natural sciences and earth sciences. It states that systems are complex and dynamic, and that subsystems interact in such a way as to be unpredictable, in order that eventually, unique systems are built. While systems might resemble each other broadly (we recognise a cloud, or an oak tree), each version of it is also unique. Within the field of applied linguistics, Complexity Theory and Dynamic Systems Theory share assumptions that language learning is a complex, dynamic, adaptive system which varies from individual to individual. The label CDST (Complex Dynamics Systems Theory) is a unifying label which brings together the work of Larsen-Freeman (Complexity Theory: e.g. Larsen-Freeman & Cameron, 2008a; Larsen-Freeman, 2006), and that of the Groningen Group – De Bot, Lowie, Verspoor and Van Geert (Dynamic Systems Theory e.g. De Bot, Lowie, & Verspoor, 2007; Lowie, 2013; Ortega & Han, 2017).

CDST is a metatheory (MacIntyre, MacKay, Ross, & Abel, 2017), defining not only the nature of the object, but also the methodology with which the object is to be investigated (Larsen-Freeman, 2017). It explores several variables in motion, relative to one another, and in the context in which they operate (Larsen-Freeman & Cameron, 2008), meaning that a longitudinal perspective with multiple data collection points is necessary.

The ‘object’ for the present study – listening proficiency, with all contributory elements – is illustrated in Figure 4.

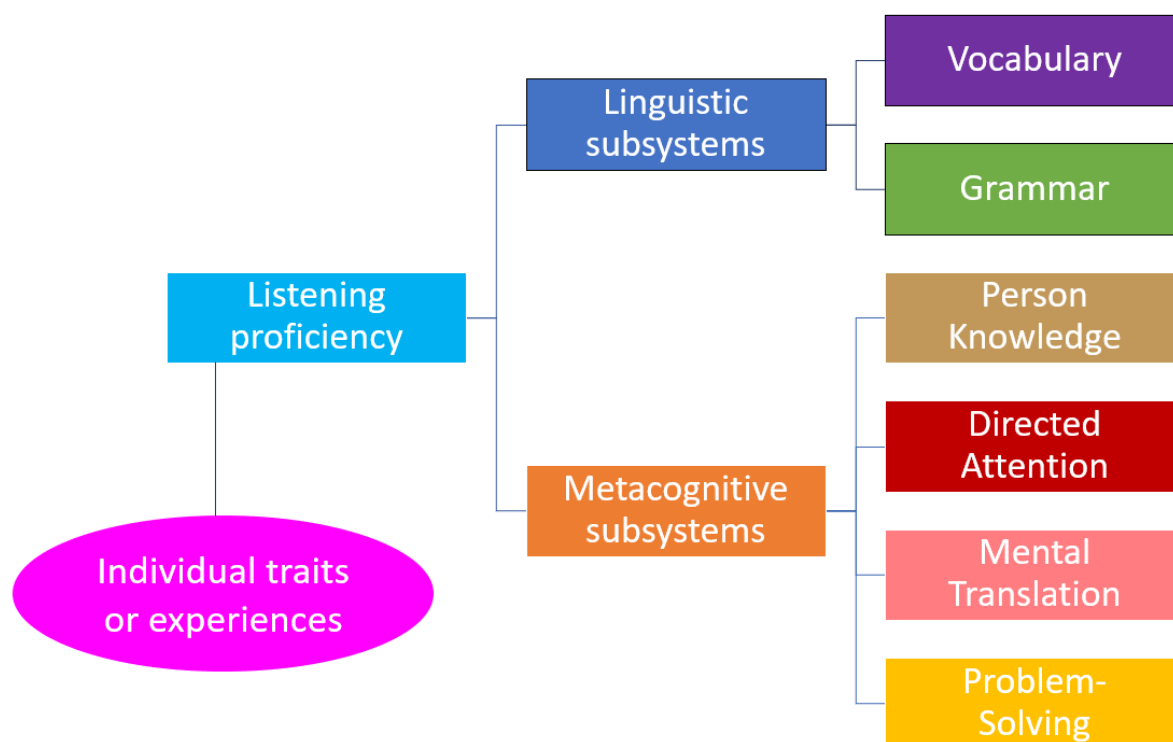


Figure 4: The proposed complex dynamic system of the present study

The starting point for the choice of subsystems was with Field’s model (2008b, 2013) (Figure 1). Although there is evidence that they contribute to L2 listening (Long, 1990), the factors ‘external knowledge’ and ‘current topics’ were ruled out before piloting due to the difficulty of creating valid and reliable tests (Vandergrift & Baker, 2015), particularly for beginners within a longitudinal framework. Other top-down elements are conceptualised through the model offered by the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006) (see appendix 15.1.1 for full details). Working Memory was piloted but ruled out (see appendix 15.1.3 for full account). It is worth acknowledging that beyond the dynamics of the learner’s cognitive context (in the case of the present study, their listening skills), other factors which might be important within the CDST model could include a social context (on the right of the diagram), and a

pedagogical context (on the left of the diagram), and each of these contexts could also be said to be complex dynamic systems, of which the researcher needs to be aware (Larsen-Freeman & Cameron, 2008). Taking a case-study approach will enable me to explore other factors which might be key to specific individuals.

2.6.1. The advantages of a CDST approach

Given that the present study aims to track the developmental trajectory of L2 listening, and the relationships between different listening subsystems throughout the process, CDST appears to be a natural fit. Indeed, Ortega and Han (2017) state that among the broad requirements expected of CDST research are the explicit inclusion of time and change, including multiple waves of data collection; the capture of continuously interacting subsystems across time; and the measurement of individual variability. The present study speaks to the need for more longitudinal work on second language development (Hulstijn, 2015), including that which includes dense data bases ‘to enhance our understanding of the fine-grained patterns of change over time’ (De Bot, Lowie, & Verspoor, 2007:7), in order to capture its ebbs and flows in detail, and the range of factors which influence this dynamism.

2.6.2. A match with theoretical frameworks of listening?

Not only does CDST appear to be an appropriate metatheory given the proposed research design of the present study, but it also fits neatly with both theoretical frameworks of how listening works, and the current literature pertaining to L2 listening.

Field’s model of listening (Figure 1) (2008b, 2013) bears many hallmarks of a complex dynamic system. It has a range of subsystems which are constantly interacting, the balance between which is changing depending on the nature of the input as the input

changes. A listener's representation of what they hear is iterative, dynamic, and dependent on external factors varying from context to environment. One might argue that a listening model born of the Fuzzy Lexical Representation Hypothesis (Figure 2) as well as that stemming from the learning to listen / listening to learn model (**Error! Reference source not found.**) also have the potential to be transposed onto the CDST framework.

However, much of the current literature in the field of L2 listening aims to examine the relationship between one subsystem and 'the whole' (usually seen as success in listening comprehension, operationalised by comprehension questions or by free recall). By contrast, a CDST approach to listening takes a more holistic view in order to understand more about the constantly changing relationship between the subsystems as well as broader factors such as the listener's self-efficacy and environment (De Bot et al., 2007).

2.6.3. Precursors and dependents

Within the CDST literature, discussion exists regarding the behaviour of the subsystems in relation to one another. A key relationship is that of the precursor and the dependent: for example in a first language, vocabulary must achieve a certain level before productive grammar appears (Caspi & Lowie, 2013): the vocabulary is the precursor and the grammar is the dependent. This could be likened to the idea that a plant needs a certain number of leaves (precursors) before it can start to grow a flower (dependent) – see Figure 5. Once a specific stage in growth has been reached, the precursors support the development of the dependents (more leaves are needed to allow more flowers to form), however throughout development, dependents are also competing for resources with the precursors (both benefit from water, light, nutrients et cetera).



Figure 5: Second language development as plant growth

Based on the findings of previous research, one might hypothesise that any of the subsystems listed in Figure 1 or indeed in Figure 4 might reasonably be assumed to be precursors for listening comprehension. In other words, it might be that, say, no development in listening proficiency can take place until a certain level of vocabulary knowledge has been reached, and that vocabulary knowledge might also need to reach a

certain size before grammatical knowledge can flourish. One might also logically hypothesise that Directed Attention could also be a precursor to the development of listening proficiency, or that either Mental Translation or Person Knowledge (i.e. anxiety) might detract from the growth in listening proficiency.

What cannot be gleaned from previous research is the relative importance of each of these at different stages in the learning process, or whether different sizes of the relative subsystems results in a different overall whole – much like asking to what extent the plant with plenty of light with little fertiliser resembles the plant grown in darker conditions yet very fertile soil.

2.6.4. Complexity as messy yet valid

It is a familiar idea to a teacher that a learner is a complex being with a range of experiences and needs which vary from day to day. Yet research with intact classrooms is a ‘particularly complex and multifaceted endeavour’ (Mackey & Gass, 2005: 212). To make sense of the data that is generated in such a setting, one could attempt to ‘tame’ it by taking a positivist approach, reducing it to a small number of ‘tidy’ variables, in order for the researcher to take control. However, in the classroom context, that control is an illusion (Simpson & Rose, 2020); instead, what results is an oversimplification of the learning. As my frontispiece states, ‘precision is not truth’.

CDST is one way in which the data generated from classroom research can be valued in all its colour and mess and repeated fluctuations. Teachers recognise that ‘learners are complicated people, who bring their day-to-day experiences to into the classroom, whether they are helpful to learning or not, and that therefore all manner of factors might influence the extent to which progress in learning is made’ (Simpson & Rose, 2020:139).

The CDST approach, then, creates a shared space between practitioner and researcher (Mercer, 2016; De Bot et al., 2007): practising teachers acknowledge the complexity of interactions and processes within their learners and their classrooms (Kostoulas, Stelma, Mercer, Cameron, & Dawson, 2017). Instead of decontextualising learning, the approach sits comfortably with the many complex processes and events that happen in the classroom, and as such, one might argue that it is a more ecologically valid approach to classroom research than more ‘traditional’ positivist methods. These ideas will be further explored in Chapter 3, Methodology.

2.6.5. The challenges associated with CDST research

The boundary issue

CDST states that there are interconnected systems which might sit alongside our primary area of interest: this can be seen in Figure 4, where the primary area of interest is the cognitive context of L2 listening, but secondary areas of interest are contexts specific to the individual, which might also be social, affective or pedagogical. These interconnected systems might influence the primary area of interest in unpredictable ways, which begs the question as to where boundaries are drawn: how does the researcher choose what is important and what to disregard?

In the present study, the focal area is the relationships within the cognitive context, yet, consistent with the CDST approach, retains reference to the social and pedagogical contexts in which the learning is taking place (Larsen-Freeman and Cameron, 2008). At the stage of data analysis, this will allow me to seek explanations for findings in the data that will be outside the primary area of focus.

The noise issue

At the data analysis stage of a CDST study, ‘noise’ (defined as non-linear progress or unexpected outliers) in the data is not necessarily as measurement error, but instead as ‘sound’ (Marin & Peltzer-Karpf, 2009), to be observed and interpreted where appropriate. This is due to the fact that CDST is a theory of process rather than product (Caspi & Lowie, 2013): if one accepts that underlying development is the interaction of a range of subsystems, it could be that unexpected noise in the data might be indicative of a change in the system – for example a change in the relationships between the precursors and dependents mentioned in section 2.6.3. However, for this to be the case, it is essential that the researcher is satisfied that instruments used are valid and reliable measures of the variables under examination. This will be further unpacked in chapter 3.

2.6.6. Summary

Within the context of the present study, CDST has much to offer. Given that the present study is situated in the unique environment of a single school, it will be important to resist claiming that any findings are generalisable, but similarly, this conventional view of a lack of generalisability does not minimise the usefulness or relevance of the work: ‘Case-driven does not equate with ‘idiosyncratic’ and ‘complexity’ does not equate with ‘randomness’ (Sealey & Carter, 2004:210).

2.7. Research questions

Having reviewed the literature, and conceptualised L2 listening progress as a complex, dynamic system, I pose the following questions:

In the context of instructed beginner learners of French in a UK secondary school:

1. What do listening errors of school-aged students of French reveal about a) their listening processes, and b) the development of their listening proficiency over a three-year period?
2. How do the relationships between listening proficiency and subsystems of a) language knowledge (vocabulary and grammatical knowledge and segmentation skills) and b) metacognition (Person Knowledge, Problem-Solving, Mental Translation and Directed Attention) develop over three academic years?
3. To what extent, and how, do the answers to the above questions vary between individuals, and how do individual patterns compare to those seen at group level? Are there other factors in the students' experience which are relevant to their listening proficiency development?

3. Methodology

This chapter outlines the methodology of the project. It begins by setting out my own theoretical and ontological position, and the details of the research design. Then it aims to set the design into context by discussing the creation and piloting of instruments in more detail. The procedures surrounding data collection and analysis are addressed. I conclude with ethical considerations, problems arising, including the impact of the pandemic, and methodological limitations and implications.

The research makes a unique contribution to second language listening through its long period of data collection, its view of listening from three perspectives, and its use of instruments to uncover new perspectives on the experiences of the beginner listener. At each time point, the 100 or so participants (80 of whose data was retained after an exclusion process) completed a test battery which included listening proficiency assessed through paused translation, aural vocabulary and grammaticality judgement tests, and strategy and self-efficacy questionnaires. The data generated allowed me to examine the complex and dynamic relationships at play during listening. To recap, the research questions are:

In the context of instructed beginner learners of French in a UK secondary school:

1. What do listening errors of school-aged students of French reveal about a) their listening processes, and b) the development of their listening proficiency over a three-year period?
2. How do the relationships between listening proficiency and subsystems of a) language knowledge (vocabulary and grammatical knowledge and segmentation skills) and b) metacognition (Person Knowledge, Problem-Solving, Mental Translation and Directed Attention) develop over three academic years?

3. To what extent, and how, do the answers to the above questions vary between individuals, and how do individual patterns compare to those seen at group level? Are there other factors in the students' experience which are relevant to their listening proficiency development?

3.1. Theoretical and ontological orientation

CDST rejects notions of linear causality, and is a metatheory (Larsen-Freeman & Cameron, 2008) which challenges the 'conventional expectation that a good theory is one that describes, explains and predicts' (De Bot & Larsen-Freeman, 2011:23). My own perspective additionally aligns with those who view Complexity as an ontology, (e.g. Knight, 2022; Osberg & Biesta, 2010), which leads in turn to a specific epistemology and methodology. In taking a complexity approach which is congruent with my extensive teaching experience, I hope to offer a perspective on the real world of language learning and teaching (Rose, 2019) and attempt to account for the fact that my participants' lives outside the classroom contribute to the learning which takes place within it (Ushoida, 2021).

In its proposal of a transdisciplinary framework for SLA research, the Douglas Fir Group (2016) urges the researcher to 'think integratively' (p38). My practice as a researcher is firmly rooted in my practice and experience as a classroom teacher: 12 years teaching French, German and Spanish to English secondary school students, and ten years preceding that teaching small groups of adult learners. It is an experience that leads to me, for the present study, to reject positivism, with its insistence on objectivity which cannot be adequately adapted to the messy experience of real people with real lives (Rose, 2019; Simpson & Rose, 2020). An objectivity which 'erases our humanness and that of our participants' (Consoli & Ganassin, 2022:4; Consoli, 2022) does not sit right

with me, particularly in the case of the present longitudinal study over three years where one cannot erase or discount the richness of the very experience of the research, of getting to know the participants and their lives.

I also question the extent to which it is reasonable to claim that L2 listening can be researched from a positivist perspective, if positivism claims that all reality is objective and external to the mind (McKinley & Rose, 2019), and knowledge is reliably based on observed objects and events. The reality of listening is *internal* to the mind of the listener, meaning any data on the listener's comprehension are proxy data in some way or another. By extension, the very choice of instrumentation by the researcher is subjective to some degree or another, detracting from its positivism.

As such I propose that, in undertaking this research, another complex dynamic system has come into play: one whereby such issues as ontology, epistemology, axiology, positionality, and even utility interact with my own views and experiences during the developmental trajectory of this research and of me as a researcher, and, crucially, a teacher-researcher. Indeed it is my position as a teacher-researcher, and my deep and continued reflections on my experiences of teaching and learning, which have led me to beliefs about what is valuable, and this informs my own stance on both what is true, and in turn, how we know what is true. I recognise that within a positivist approach to research, this insider position could be seen as a limitation or an indication of bias, but instead I aim for transparency in my relationship with the research in order to 'identify, understand, capitalise on and celebrate its impact' (Consoli & Ganassin, 2022:3).

3.1.1. Positionality

My understanding of what I was observing in school, even the quantitative data, was inevitably coloured by my experiences of being a teacher in that school, and it would not

have been possible to extricate myself from this kind of prior knowledge (McKinley, 2017) in order to establish some positivist ‘objectivity’. I acknowledge that, in theory, I could have sought out another school for data collection; see section 12.1.5 in the Limitations chapter for more discussion on this. Even if a different research site had been feasible, I assert that my long experience as a teacher of languages would have continued to influence the interpretations which I made of the research experience.

Czerniawski (2022:14), in his research into the positionalities of teacher-researchers carrying out doctoral projects, recognises the ‘degrees of ambiguity, fluidity and uncertainty’ inherent in this dual role. I have already stated that I was a teacher-researcher and that the project took place within my own school. My piloting took place when I had been employed there for an academic year, and main phase data collection happened between the start of my third year and the end of my fifth. However, the insider-outsider dichotomy (Akkerman & Bakker, 2011) is not entirely straightforward: for example, I expressly asked not to teach the classes which were the subject of my research, because I wanted to ensure that the research remained observational and did not accidentally tip into an intervention as my pedagogical instincts responded to interim findings I was making. This distance from the research classes was mostly successful, although I had contact with the participants in other contexts in school. This personal history is inevitable in the context of the daily life of a teacher (Caelli, Ray, & Mill, 2003) and then by extension, a teacher-researcher.

As a doctoral-level teacher-researcher, therefore, my positionality was influenced by the dual roles I inhabited, which reflected both the opportunities and ethical complexities highlighted by Consoli (2021) and Galloway (2016). Navigating the intersection between educator and researcher is unusual at this level, and I found myself in a position where the additional knowledge of my students’ learning contexts could enrich my research

insights at a deeper level than had I been a visiting researcher. However, this proximity also presented challenges, particularly regarding power dynamics and potential biases. Consoli (2021a) emphasizes that the dual role can lead to incidents requiring ethical reflexivity to manage the blurred boundaries, and I wrote about them in my reflexive diary. Galloway (2016) discusses the difficulty with ensuring that students feel free to participate in research without pressure. An example of this took place at Time 4 when I saw case study Fran in the corridor, and she told me she did not want to do the listening survey that day. (I told her as a teacher she needed to come to the timetabled room and then we'd have a chat about how she felt about participating; she decided to do so.) As such, my positionality required me to be mindful of how my position as a teacher could influence students' responses, and to adopt strategies addressing potential ethical risks, such as transparent communication and voluntary participation. Embracing a reflexive approach, I aimed to remain aware of how my dual role shaped the research process.

For example, my professional role in the school allowed me to consider my research on a daily basis as I went about my teacherly work. For the student participants, I was aware that it was likely they would see me more as a teacher than as a doctoral researcher: they saw me on a daily basis wearing a 'staff' lanyard, and I had regular interactions with their friends and siblings as a teacher in the school. In this context I needed to acknowledge that there would be incidences in which they might be guarded in what they told me as a researcher, as I was first and foremost a teacher to them.

My position will have influenced the way in which I interpret my qualitative data, too, due to the relationships I have built up with the participants, and my insider knowledge and understanding of the learning context. It was unpredictable how I might interact with a participant: a corridor interaction, attendance at a detention, a lesson I might be covering.

One very practical conclusion of this positionality statement is my conscious choice of the use of the first person for the remainder of the thesis, and a deliberate avoidance of both the passive voice and reference to myself as ‘the researcher’. (Banegas & Consoli, 2024:7) encourage such a reflexively transparent approach which runs counter to ‘protecting notions of methodological perfection’. Instead, I acknowledge my position as the researcher and the colour and bias that my own experiences – ‘life capital’ (Consoli, 2022) – bring to the project.

3.2. Overview of research design

3.2.1. Aims and research questions

The present study aimed to understand the developmental trajectory of second language listening of French in a limited input instructed classroom context. Table 2 sets out the research questions and the types of data analysis that were used to address them.

Table 2: Summary of research questions, sample, instruments, and data analysis

Research question	Sample	Instruments	Data analysis	
In the context of instructed beginner learners of French in a UK secondary school...				
1	What do the listening errors of school-aged students of French reveal about a) their listening processes and b) the development of their listening proficiency over three years?	Cohort of 80 students of French at the participating school from September 2018 to July 2021	Three paused audio translation tests at each of 20 time points	Qualitative data analysis through reflexive thematic analysis
2	How do the relationships between listening proficiency and the subsystems of linguistic knowledge (vocabulary and grammar) and metacognition (Directed Attention, Person Knowledge, Problem-Solving and Mental Translation) develop over three academic years?	Cohort of 80 students of French at the participating school from September 2018 to July 2021	At each of 20 time points: audio paused translation tests, AuralLex vocabulary test, aural grammaticality judgement test, modified Metacognitive Listening Awareness Questionnaire	Descriptive statistics analysed through CDST approaches (moving min-max, moving correlations)
3	To what extent, and how, do the answers to the above questions vary between individuals, and how do individual patterns compare to those seen at group level? Are there other factors which are relevant to their listening proficiency development?	A subset of the cohort above, consisting of four students who volunteered as case studies	All the above, plus semi-structured interviews and stimulated recall	Both of the above, in individual scales, for each of the four case studies

3.2.2. Population

The population is secondary-school learners of French in England. They are children who begin French at secondary school aged 11, in year seven, although many will have covered some very basic French at primary school. In general, the population learn French in an input-poor environment, in that they will have around two and a half hours' teaching a week during school term time, which amounts to a little less than 100 hours' learning in the school year (British Council, 2023), although each school will vary in respect of the number of hours dedicated to French. Outside the classroom, a small amount of homework will be set – up to about half an hour a week (therefore a total per academic year of perhaps 20 hours), although for practical reasons it is almost unheard-of that this would involve listening to French.

Learning a foreign language is compulsory for the first three years of secondary school (Department for Education, 2023), i.e. from year seven to year nine, which is the academic year in which pupils turn 14 years old. French is the commonest foreign language taught (Board & Tinsley, 2016). From year 10, languages are optional, with about half of the cohort nationally choosing one to study for their General Certificate of Secondary Education (GCSE), the national examination taken at age 15-16. In 2023, 126,560 pupils took a GCSE in French (to contextualise: 786,815 took mathematics in the same period) (Ofqual, 2023): each year since 1993, fewer candidates have taken French (Stubbs, 2017). Concern is regularly expressed that language-learning in England is declining (Lanvers & Coleman, 2017).

3.2.3. Institutional context

The data were collected in a single secondary school in a city of 150,000 people in England – one of seven secondary schools in the city. The school has around 1200

students, from age 11 (year 7) to 18 (year 13). It is rated ‘good’ by Ofsted. Table 3 places the school into the English context (Department for Education, 2019).

Table 3: National context of study setting

	Study setting	Schools in England average
Size of school	~1200	956
Absence rate	6.6%	5.5%
Special educational needs	11.6%	10.8%
Speakers of English as an additional language	13.2%	16.9%
Eligible for free school meals	14.2%	27.7%
Study a language at GCSE	38%	40%
Achieve grade 5 in English and Maths	69%	43%

The school was chosen out of convenience, as it had been my employer for two years at the start of data collection. The practical demands of the research design, with its 20 data collection points, 80 computer room bookings, approximately 160 separate interviews over a three-year period meant that such research would be very difficult to undertake as an external researcher.

3.2.4. Participants

As the new cohort began at the participating school in September 2018 (year 7), the 103 students who were learning French were invited to participate (through opt-out rather than opt-in procedures: see Section 3.7). The students had a range of profiles and backgrounds; 31 reported that they are regularly exposed to a language other than English outside school. These languages are set out in Table 4. Eighteen of these participants claim to use the other language regularly.

Table 4: Languages to which participants are exposed outside of school.

Language	Number of participants
Arabic	6
Bulgarian	1
Danish	1
Filipino / Tagalog	2
French	1
Hebrew	1
Italian	2
Kurdish	1
Mandinka	1
Portuguese	4
Rwandese	1
Shona	1
Slovakian	1
Spanish	2
Swahili	3
Urdu	3
Wolof	1
Total	32

Four students were also selected as case studies, and were interviewed immediately after each test battery; all those who were shortlisted had been categorised by the school as monolingual English speakers, which had been my initial plan in order to reduce confounding variables. However, the background information given in Table 5 shows that the school data were not correct, which I did not discover until my first interview with each at Time 1, at which point I made the decision to continue with the case studies I had, and view Theo's and Fran's additional language knowledge as an asset. Prior attainment level is a measure used by the school and was not collected specifically for the present study.

Table 5: Details of case studies

Pseudonym	Gender	Prior Attainment	Home languages
Helen	F	Higher	English
Theo	M	Higher	Mother speaks a mix of Swahili and English at home; Theo occasionally uses Swahili.
Fran	F	Middle	Italian is Fran's language of choice at home, although English is also spoken.
Ryan	M	Middle	English

3.2.5. Testing schedule

In the English secondary school system, the academic year lasts 39 weeks and is split into six short terms, each lasting between five and eight weeks. Terms offered between 15 and 24 hours' worth of input in French, plus any homework. Testing took place at the end of each term, with two additional tests at the start of the new academic year in year 8 and year 9, to give a total of 20 time points. This schedule is illustrated in Table 6. While it is acknowledged that this does not make entirely equally spaced tests, it better reflects the curriculum which tends to fit around term dates. It is important that a testing schedule is frequent enough to capture the aspects which underlie changes in the learners' abilities (Larsen-Freeman, 2006). Piloting (see section 3.4) confirmed that this testing schedule was appropriate.

Table 6. Test dates and term dates for first year

Time	Term	Term Dates	Date of test
1	1	5 th Sept to 19 th Oct 2018	w/c 15 th Oct 2018
2	2	29 th Oct to 21 st Dec 2018	w/c 17 th Dec 2018
3	3	7 th Jan to 15 th Feb 2019	w/c 11 th Feb 2019
4	4	25 th Feb to 5 th Apr 2019	w/c 1 st Apr 2019
5	5	23 rd Apr to 24 th May 2019	w/c 20 th May 2019
6	6	3 rd June to 24 th July 2019	w/c 15 th July 2019
7	1	4 th Sept to 25 th Oct 2019	w/c 4 th Sept 2019
8			w/c 21 st Oct 2019
9	2	4 th Nov to 20 th Dec 2019	w/c 16 th Dec 2019
10	3	6 th Jan to 14 th Feb 2020	w/c 11 th Feb 2020
11	4	24 th Feb to 3 rd Apr 2020	w/c 16 th Mar 2020 ²
12	5	20 th Apr to 22 nd May 2020	w/c 18 th May 2020
13	6	1 st June to 17 th July 2020	w/c 6 th July 2020
14	1	3 rd Sept to 23 rd Oct 2020	w/c 5 th Sept 2020
15			w/c 19 th Oct 2020
16	2	2 nd Nov to 18 th Dec 2020	w/c 14 th Dec 2020
17	3	5 th Jan to 12 th Feb 2021	w/c 1 st Feb 2021
18	4	22 nd Feb to 31 st Mar 2021	w/c 22 nd Mar 2021
19	5	19 th Apr to 28 th May 2021	w/c 24 th May 2021
20	6	7 th June to 21 st July 2021	w/c 12 th July 2021

² testing brought forward anticipating first covid-19 lockdown

3.2.6. Locations and equipment

At the beginning of data collection, it was anticipated that at each time point, each class would have a single French lesson timetabled into a computer room in the last week of term, during which some of the lesson time would be devoted to data collection. Data was collected through web-based surveys on the Qualtrics platform. Until February 2020 (time 10), and from April to July 2021 (times 18 to 20) data were collected at school through desktop computers, with students listening to the audio files on RP-HT22 Panasonic headphones. Due to covid, Time 11 was hybrid, with two classes completing the surveys from school, and the other two from home. From Time 12 to Time 17, participants carried out the surveys at home on whatever devices they had available; see section 3.8.1 on the impact of the pandemic for more information.

3.3. Methodological challenges and possible solutions

This section will explore the principal challenges encountered when designing research into the complex, dynamic systems surrounding listening, and discuss solutions to these challenges made by researchers, and concludes with my own solution for each challenge. In this way it paves the way for the discussion of piloting my solutions. It divides the challenges into three subsections: those concerning L2 listening; longitudinal challenges, and CDST challenges. It draws on a range of research, and in particular will return frequently to the few previous papers on listening from a CDST perspective.

3.3.1. Operationalising understanding spoken language

Defining listening and measuring comprehension is the cornerstone of the present study. This section, then, will address ways in which listening proficiency can be

operationalised, and to explore the extent to which these methods might be compatible with CDST research designs.

Because listening takes place in the mind, and is fleeting, various proxies are used in order to gather data about the level of understanding that has taken place. The most common proxy is a question and answer (q&a) test in response to a spoken text. Field (2019:79) lists a variety of options here: ‘multiple choice, visual multiple choice, gap fill, form-fill, multiple matching, labelling, drawing lines to link graphics’, all of which are easy to mark. Yet no q&a approach is comprehensive: the resource creator is deliberately tapping into a specific element of understanding of the input and deliberately overlooking others. The approach also makes varying demands on other skills such as reading proficiency, strategy use or working memory (Field, 2019).

I was not keen on the q&a, which, I felt, lacked both nuance and rigour. The data that it would generate would be highly dependent on the nature of the questions posed, and without stimulated recall for every participant, I would not know whether the answer given was a guess or based on (mis)understanding of the input. An alternative to the q&a test is a paused transcription, or paused translation exercise (Field, 2008b; Yeldham, 2017b), which removes both these confounds. In this kind of task, an audio track is paused, and the listener is required to write what they have heard. This could be in its entirety, or a certain number of words (Estes, 2014).

The listener can write what they have heard either in their L1, or in the target language, or both. While (Field, 2008b) advocates for same-language transcription, he acknowledges that this begins to tap into an assessment of L2 writing proficiency.

Listeners with good working memory are also at an advantage given that they could simply reproduce what is in their echoic memory, and as such it does not provide the

researcher with data on how the language is being processed or understood. It was for these reasons that I chose paused translation as my principal instrument, which removes the L2 writing proficiency and working memory confounds, while providing me with rich data.

3.3.2. Longitudinal challenges and solutions

Collecting repeated measurements over a long period of time presents a unique set of challenges, in particular when they intersect with L2 listening tests.

Repeated measures and the challenge of the test effect

The challenge of the test effect might be one reason why listening (and reading) have been so overlooked by CDST researchers. When exploring other linguistic skills such as writing and speaking, researchers can give the same prompt many times over, yet doing the same thing for reading or listening will create a ‘practice effect’, whereby the mere act of repeating the test is likely to bring about an improvement in scores (Rogers & Révész, 2020). Therefore the CDST researcher into listening needs to source or create many similar tests in order to measure their learners’ developmental trajectory.

In this regard, the use of a standardised q&a test offers some advantage to the CDST researcher. Firstly, standardised tests such as those used by both Dong (2016), who chose the Chinese ‘College English Test’ and Chang (2017; Chang & Zhang, 2021), who used IELTS and the College English Test, are well established and tested, with many iterations (Yeldham, 2017b) of equal difficulty. With data collection points typically numbering into double figures, this is an advantage. Unfortunately, there is no universally accepted French language equivalent – even the *diplôme d’études en langue française* (DELF) only offers six tests at each level.

The pros and cons of the transcription approach discussed in section 3.3.1 are almost the opposite to those of the standardised q&a test, in that the researcher can easily source material. This means that within the CDST context, large numbers of tasks can be created relatively easily for use as repeated measures.

The challenge of participants' progress during the study

In the cases of both Dong (2016) and Chang (2017), the assumption was made by the researchers that instruments of equal level could be given to participants throughout their projects. However, in my own research setting, progress was expected during the course of the data collection period, necessitating the use of instruments which would continue to grow with the proficiency of the participants. Again, the circumstances pointed to the appropriateness of paused translation, at a difficulty level which matched that of the participants. See section 3.4.1 on my choice of listening passages which sets out how I chose passages to minimise floor and ceiling effects while the participants continued to progress.

Transforming the results of multiple tests onto a common scale

Tracking learners' development of receptive skills over an extended period of time brings with it the challenge of having to match the difficulty of the instruments with the proficiency of the learner, in order to avoid floor and ceiling effects in the tests. Again, had I been able to access the DELF, I could have underpinned my data with an assumption (or verification) that the test-makers had already done this. With self-made tests, as tests increased in difficulty over a timeframe, the raw data needed to be transformed to a common scale.

I tackled this problem through use of a Rasch model, which allowed a universal calibration of different tests. Different tests taken at different times can be placed on a

common scale in order that development can still be tracked (Boone, Staver, & Yale, 2014; Boone & Staver, 2020; Meyer & Zhu, 2013). To facilitate this, there must be some repetition of items between time points – these repeated items create anchor points that allow tests to be compared. In this way, we can compare a participant who performs highly on one test with a student who performs less well on a different test. It also allows for missing data, which is inevitable when we are working with multiple data collection points (Luo, Seow, & Chin, 2001). Recommendations about sample size when using a Rasch model depend on the ability of the test items to discriminate between individuals, but a model with 30 participants and 30 items is a feasible minimum (Linacre, 2012; Wright & Tennant, 1996).

One argument sometimes cited against using such models is that of Rasch's need for unidimensionality of construct. L2 listening – and indeed many constructs within second language development – are made up of many overlapping constructs and could not be said to be unidimensional. Furthermore, inevitably data collection instruments tap into skills beyond the construct (or dimension) to be measured. However, Fan & Bond (2019) argue that this is a misunderstanding of the concept of dimensionality, which can be interpreted both psychologically and psychometrically. In other words, the Rasch model accepts that many psychological processes might be necessary in order to succeed on a test item, but what is important is that these processes 'function in unison – that is, performance on each item is affected by the same processes and in the same form.' (p84)

Although Rasch modelling might then appear to be a neat solution to the problem of tests of varying difficulty when collecting CDST data, another similar approach is the use of the Item Response Theory model (IRT). With Rasch, data is adapted (outliers are removed) in order to fit the model, and with IRT, outliers are not removed (Boone et al., 2014). It could be argued, therefore, that IRT is more compatible with the exploratory

nature of much CDST research whereby all data are considered valid and would not be removed for failure to fit a model. While this is an important detail, however, Andrich & Marais (2019) assert that the difference between the two is philosophical only, and furthermore, IRT models also demand the kinds of sample sizes that are likely to be incompatible with classroom research (Meyer & Zhu, 2013).

3.3.3. CDST and the challenge of the boundary

CDST states that there are interconnected systems which might sit alongside our primary area of interest. Interconnected systems influenced my primary area of interest in unpredictable ways that I had not initially anticipated, resulting in three complex systems on three different scales – the moment-to-moment experience of listening (Chapter 4); a cognitive and metacognitive perspective over three years (Chapters 4 and 5) and an individual perspective, which allowed me to explore the presence of further variables which were specific to each case study (Chapter 6). This collection of systems illustrates the way in which CDST researchers grapple with the question as to where and how boundaries are to be drawn (Larsen-Freeman and Cameron, 2008). It is these underpinning principles of the CDST approach which justify the nature of the research questions. The interconnectedness of a range of subsystems is recognised, but the challenge presented to research is also accepted (De Bot & Larsen-Freeman, 2011). It is standpoints such as these which lead me to see Complexity as an overarching *ontology*, rather than a metatheory.

3.4. Piloting instruments

A full and detailed account of the creation and piloting of instruments is available in appendix 15.1. This section is a summary of the instruments' creation and pilot.

3.4.1. Choice of listening passages

In order that the measures chosen developed along with the learning of the students over the three years of the study, and align with their scheme of work, listening comprehension texts (examples in appendix 15.3) were carefully chosen from the syllabus, and graded according to difficulty. Initially, listening texts were sorted by grade assigned by the textbook. Then I double-checked the grades by assessing the speed of delivery in words per minute, the lexical richness of the text by noting the percentage of words in each frequency band, according to lextutor.ca. The difficulty of the passages was also piloted, whereby students in years 7, 8 and 9, and teachers of French, listened to each text and rated difficulty on a five-point Likert scale.

Field (2008b) suggests undertaking a detailed analysis of texts for listening assessment which would include not only lexical richness, but also syntactic complexity, denseness of ideas, degrees of repetition and redundancy. This was considered but ruled out given the very short and simple nature of the texts.

As for piloting the appropriacy of timing, three pilot test batteries were held with the previous year seven group, at the end of term five and half-way through term six to ensure that the testing schedule was frequent enough to capture some level of progress. I concluded that this timing was indeed appropriate and did not need to be made more frequent.

3.4.2. Choice of response to listening passage

Section 3.3.1 discussed the differing ways in which researchers assess listening proficiency for research purposes. For the present study, to measure comprehension a number of different question types were considered: open questions; multiple choice

questions; and free recall with a mark-scheme based on how many idea units a participant correctly identified.

3.4.3. Piloting the listening passages

It was important to select a range of listening texts whose difficulty would increase in line with the participants' growing French proficiency over the three years of the study. 160 listening passages from GCSE past papers were assessed for suitability: each was coded according to the topic, the percentage of words in each 1000-frequency level, up to K20, according to lextutor.ca, the number of words and the duration, which also gave a score for speed of delivery. A screenshot taken from part of the spreadsheet is given in Table 7.

Table 7: Paused translation pilot: spreadsheet to find appropriate passages

Code	sub topic	title	notes	word	dur	dura	wpm	K	off	K1	K2	K3	K4	K5	K6	K7	K8	K9	
C41	eating out in	a quick snack		112	80	1.33	84.00	1.68	79.83	82.4	86.6	87.4	89.9				91.6	92.4	93.3
C51	subjects	my lessons		71	50	0.83	85.20	1.28	85.9	91	93.6				94.9			96.2	
C56	school rules	encore des règles		67	38	0.63	105.79	2	85.33	88.7	94	94.7	95.3				96	96.7	
C58	primary school	my primary school	year 10 pilot	72	40	0.67	108.00	0	84.81	89.9	94.9		96.2				97.5		
C61	school activities	what's going on in school		74	44	0.73	100.91	0	80	87.5	97.5	98.8							
C62	exchanges	exchange plans		100	56	0.93	107.14	0	88.07	90.8	96.3	98.2						99.1	
C76	training	apprenticeships		90	50	0.83	108.00	0	88.42	90.5	93.7	95.8	99				100		
C112	les rapports p	les rapports personnels		44	33	0.55	80.00	0	93.48	95.7	97.8								100
C14	music	a young singer	year 9 pilot	80	52	0.87	92.31	1.11	83.33	87.8	91.1	92.2		94.4			96.7		
C16	daily routine	Morning at home	year 8 pilot?	119	60	1.00	119.00	0	84	88.8	92	94.4	96					96.8	99.2
C17	breakfast time	what do they eat for breakfast	year 9 pilot	69	39	0.65	106.15	1.14	76.06	80.3	87.3	90.2					91.6	94.4	
C21	visitor information	a visit to the zoo	year 9 pilot	94	69	1.15	81.74	0.95	78.1	85.7	94.3	97.2	98.1						
C22	things to do	going on an outing	year 7 pilot	95	53	0.88	107.55	0	87.63	94.9	97.9							99	
C54	shopping preferences	my favourite shops		125	82	1.37	91.46	0	86.43	89.3	93.6	95.7	96.4	97.1					
C55	at the train station	at the station	year 10 pilot	128	90	1.50	85.33	0	87.5	91.7	95.1	97.2	97.9					98.6	
C60	school subjects	my lessons	year 7 pilot	71	50	0.83	85.20	1.28	85.9	91	93.6				94.9			96.2	
C62a	school routines	school		83	56	0.93	88.93	0	78.89	84.5	92.2	94.5	96.7				98.9		
C66	future plans	school and beyond		70	48	0.80	87.50	0	82.43	91.9	95.9		97.3	98.6	100				
C76a	my work experience	my work experience		50	34	0.57	88.24	0	89.29	91.1	96.4	98.2	100						
C2	pets	opinions of pets		90	60	1.00	90.00	0.1	84.47	92.2	95.2	96.1							
C11	eating in a cafe	at the cafe	year 7 pilot	59	52	0.87	68.08	0	81.82	85.5			92.7				94.6		
C17a	social issues	jobs that help others	year 9 pilot	54	44	0.73	73.64	0	89.83	94.9			96.6	98.3	100				
C23	television	tv programmes	year 8 pilot	118	91	1.52	77.80	1.57	81.1	89	93.7	94.5	95.3	96.1				96.9	
C28	internet professions	opinions of the internet	year 9 pilot	69	49	0.82	84.49	5.19	81.82	87	90.9	93.5	94.8						
C37	holiday destinations	holiday experiences		110	68	1.13	97.06	0	87.39	91.6	96.6		97.5	98.3					
C38	holiday destinations	how often do they stay there		81	56	0.93	86.79	0	83.72	89.5	95.3					96.5		97.7	
C48	my room	rémy's room	year 8 pilot	93	49	0.82	113.88	0	91.09	93.1	95.1	96	97	98	99	100			
C53	tourist attractions	activities in town	year 10 pilot	96	72	1.20	80.00	2.75	86.24	89.9	93.6		96.3						
C81	opinions about	job choices	year 9 pilot	76	50	0.83	91.20	0	91.57	95.2			96.4	97.6			98.8	100	
C84	my work experience	how was their work experience	year 9 pilot	116	90	1.50	77.33	0	83.85	87.7	91.6	92.3	97.7	99.2	100				

This initial data provided me with a good estimate of the difficulty levels of tests.

Combining this with knowledge of the learners' curriculum – which topics would be

studied when – let me draw up a shortlist of sixty-five tracks which might be appropriate for each year group during the main phase of the study. Groups of students from different year groups (as a proxy for proficiency levels) were asked to listen to the tracks and to rate them on a five-point Likert scale for their assessment of the extent to which they felt they understood the track, from ‘not at all’ to ‘very well’. In the cases of some of the tracks, two adjacent year groups listened.

Choosing how to present the listening texts

Participants were asked to listen to a passage once in its entirety, in order to facilitate participants’ establishment of a schema, with space to take notes if needed (see Figure 6). My supervisors and I had thought that some participants might want a space in which to take notes, but deliberately no instructions were given, to see what participants might make of the input box.

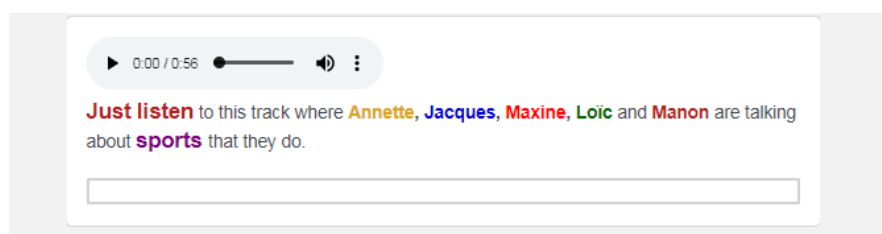


Figure 6: Pilot paused translation showing 'notes' box

In fact it was found that this caused confusion. Many pilot participants noticed it and asked me about its purpose individually during data collection, and although participants were told that it was there in case they wanted to take notes, it was hardly used. As a result, it was removed.

In order to collect further data about participants’ listening experience, they were then asked to review the transcript in order to attempt to shed light on whether the difficulties they had encountered were related to vocabulary issues or to aural processing problems.

In pilot 1, the time spent on each track in the paused translation task was recorded but not limited. At that juncture I was unsure whether participants would exploit the audio repeat functionality and listen over and over, or whether they would only allow themselves a single listen. In fact there was a wide range of approaches, which I felt meant I could not reasonably compare outcomes. I brought in a system whereby the access to each track in the passage was limited to a certain length of time (approx. the length of the track plus two seconds to find the mouse and hit ‘play’), to ensure that all had equal access.

Considering how to mark the outcomes from the listenings

Marking followed the FLEUR (Woore, Graham, Courtney, Porter, & Savory, 2018) system. In addition I spent some time experimenting with a negative mark scheme, which aimed to record what participants included in their transcriptions which had been absent from the input, but in the end this was discarded as it did not appear to be reliable or replicable.

After extensive piloting, I settled on the procedure set out in Figure 7. The ‘back button’ was disabled during this process. This meant that once completed, participants were not able to edit their response.

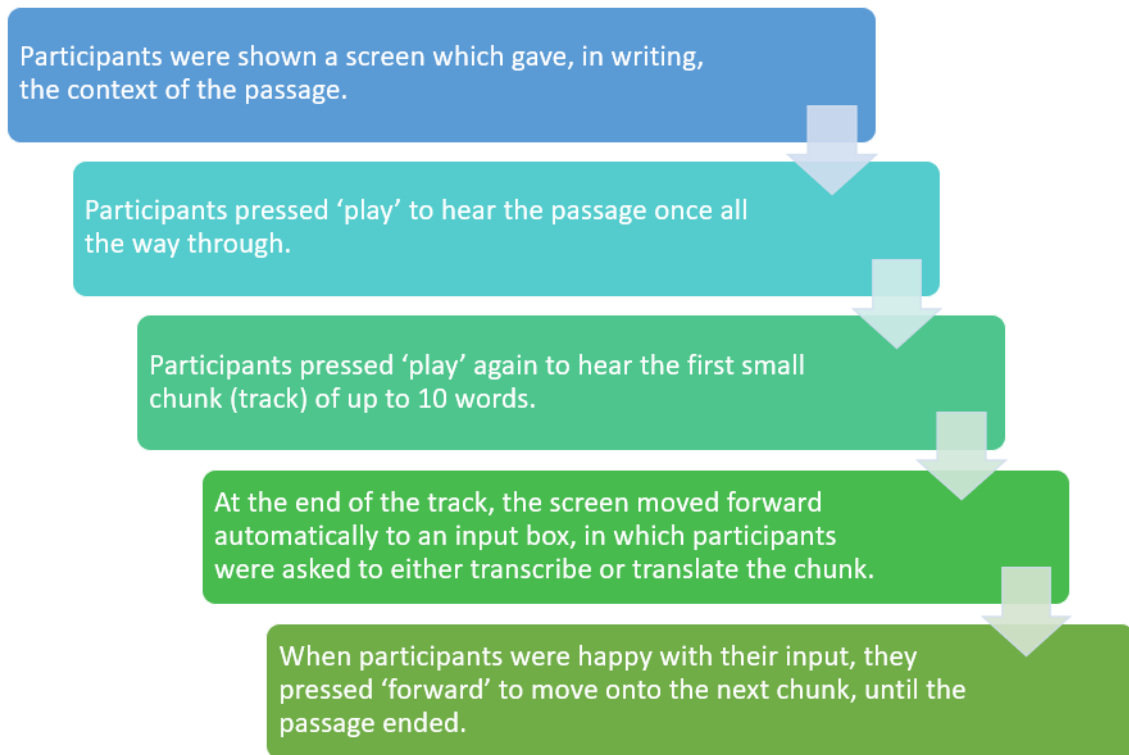


Figure 7: Paused translation data collection procedure

I began by including an optional text box for note-taking on the initial listen, given that it reduces memory constraints (Dunkel & Davis, 1994). However, on piloting I concluded that it was more distracting for participants than it was helpful, as it seemed to cause listeners to miss paying attention to subsequent sections (Alderson & Banerjee, 2002; Vogely, 1995). Not all research using paused transcription or translation allows participants to listen twice, and I acknowledge that this lacks real-world validity. However, it is standard practice in English classrooms and in GCSE examinations, so has ecological validity in terms of reflecting classroom practices.

3.4.4. Choice of scoring the paused translation passages

For the writing in English element, I needed to evaluate the data in two separate ways. For RQ1, which explored the errors made in paused translation, I used reflexive thematic analysis (Braun & Clarke, 2020; Byrne, 2021), in which error types were coded and analysed – see section 3.6.1. The second research question demanded a numerical score

for each test. Marking according to idea units (e.g. Graham & Macaro, 2008) was attempted at pilot stage, but abandoned due to the problematic nature of defining an idea unit as well as, I felt, a philosophical incompatibility with my CDST lens. Yeldham and Gruba (2016) used key words, main ideas, macro-ideas, but I struggled here with the overlap of elements and deemed the approach unreliable.

Field (2008b) recommends a simple point-per-word mark scheme, whereby each correctly transcribed word receives a point. This could then be transformed into a percentage. If the researcher is particularly interested in the minutiae of linguistic subsystems which contribute to listening proficiency, such as segmentation, vocabulary and grammar, this approach could be advantageous as it provides nuance as to the listening proficiency of the participant. If complemented with a qualitative analysis of the transcriptions, this has the potential to be highly compatible with the CDST perspective that all data are valid data, and that outliers should be considered as not as ‘noise’ to be discounted, but ‘sound’ to be explored (Marin & Peltzer-Karpf, 2009).

A more general approach would be to create a dichotomous yes-no mark scheme whereby the researcher assesses whether the gist of each sentence has been captured. Given the sheer volume of data produced by the point-per-word scheme, the dichotomous approach has the advantage of increased simplicity, but with it carries a risk of reductionism and loss of nuance.

After some trial and error, I developed a point-per-word mark scheme. I had to work through issues of contraction in French – for example, when two words in English appeared to be a single word in French (eg *j’aime* for I like, versus *je déteste* for I hate), as well as the challenges thrown up by referential cohesion (e.g. if a student translated *mes frères sont* as ‘my brother is’, ostensibly none of the words have been correctly

translated but these errors might be due to a knock-on effect of the failure to identify the plural in *mes*. The notes such as those given in appendix 15.7 are reproduced below to indicate development of my thought process:

Table 8: Developing the 'Guiding Principles' of marking the paused translations

mes frères et soeurs sont formidables			
my brothers and sisters are great	6	0	
			because they overlooked the plural. They have misheard
my brother and sister are great (or similar)	5	1	the 'mes' not the frères et soeurs'
my brother and sisters are great	5	1	overlooked the plural (it would be mon frère et mes
my brother and sister are (mistranslation of form	4	2	
my brother and sister(s) are	4	1	
my brother and sister.	3	1	points for brother and sister. Minus 1 for my as 'mes'
my brother and sisters are.	5	0	
my brother.	1	1	points for brother but not for 'mes'
my brothers and sisters.	4	0	
my dad / father	0	2	missed the correct version of 'mes' as well as the noun
my friend is	0	3	
my sister / brother is formidable	2	2	because they've missed two auditory hints at plurality: difficult because they've messed up word order. BUT
I have a brother and a sister	3	4	they have heard brother and sister so get 3 points
Other notes			large number of 'tablet' from 'formidable'

GUIDING PRINCIPLE: if in doubt, marks are about what they heard in the French rather than their English rendition, because I'm interested in their processing of the French. Therefore mes frères et soeurs - they will get marks for brother and sister because it would sound the same. The thing they got wrong is 'mes' even though that translates as 'my' because they didn't pick up on the fact that actually it's my-plural

The exploration of such possible markings resulted in a point-per-word mark-scheme which was created for each track. A typical markscheme is presented in Table 9 and illustrates the contribution of my interrater 'NRR' (a fellow languages teacher qualified to MSc level in applied linguistics), with whom I checked the minutiae of the approach.

Table 9: Example of final mark scheme for paused translation

		Accept	Details	points
1	Je joue du piano			3
2	Moi j'aime jouer de la guitare		Me, I = 1 Like to <u>play</u> / <u>playing</u> = 2 Guitar = 1	4
3	Je ne joue pas d'un instrument		I don't play = 3 An instrument = 2	5
4	Mais je suis chanteur dans un groupe	Question about article point in je suis chanteur – NRR is happy!	But = 1 I am = 2 a/ the singer = 1 In a group / band = 3	7
7	J'adore écouter de la musique classique		I love = 2 To listen to = 1 Classical music = 2	5
8	Je préfère télécharger des chansons	No point for article 'des' because we wouldn't in English	I prefer = 2 Downloading = 1 Songs = 1	4
				28

3.4.5. Measuring vocabulary knowledge

The AuralLex (Milton & Hopkins, 2005) measures receptive spoken vocabulary and is well geared to beginner learners, and its scores are better correlated with listening proficiency scores than the written X-Lex (Milton, 2009). Further, by giving a result which is an estimate of the participant's vocabulary size, it tracks aural vocabulary development efficiently. It is a well-tested instrument that already exists for L2 French and has been widely used (e.g. Milton & Hopkins, 2006; Milton & Riordan, 2006; Milton, Wade, & Hopkins, 2010).

Criticism has been levelled at the AuralLex due to the fact that it only gives information about participants' knowledge of the word's existence, and nothing deeper. However, its simple format facilitates testing for learners of low proficiency as well as young learners, and it prevents test-takers from using inference in their answers (Beglar & Nation, 2014).

However, the full French AuralLex takes 15 minutes to complete. There is only one test, which caused me concerns about panel conditioning (see section 12.1.1 in the Limitations chapter). As a result, I created ten shorter versions, each containing 48 words: 20 from the 1k band, a further 20 from the 2k band, and eight non-words following the procedure used in adapting the written X-Lex for the same population in the FLEUR project (Woore et al., 2018). In the academic year preceding data collection, I piloted these on four year groups: 7, 8, 9 and 10.

3.4.6. Piloting the AuralLex

The piloting of the vocabulary tests had three aims:

To ensure that the tests could be created online and were clear enough that participants understood the task and could complete it.

1. To pilot existing vocabulary tests and if necessary create a range of vocabulary tests of equal difficulty
2. To ensure that the tests would indicate progress by presenting appropriate challenge throughout the testing period, without having a floor or ceiling effect.

First and second pilot

I began by creating an aural version of the written X-Lex (Milton, 2006), which has previously been used successfully with year 7 students of French, and tested it on students in years 7 to 10, finding that it was both workable and showed differences in vocabulary size that increased with each year group. I then set about making a second similar test with different words in the commonest 2000 words. Comparing the results of the second test with year 7 pilot participants showed a significant difference between the two tests; I concluded therefore that they were not of equal difficulty.

Creating the GCSE-Aurallex

I identified a discrepancy between the vocabulary in Milton's work and that of the GCSE curriculum, particularly the words which appear in years 7 8 and nine. Therefore, a new list of 2000 vocabulary items was compiled, drawn from the required vocabulary list for French GCSE, as provided by the examination board AQA (AQA, 2016), in order to draw up a new set of 'GCSE aurallex' tests. I created seven tests which I hoped would be of equal difficulty, by picking 40 words at random for each list, plus eight non-words. Pilot participants took all seven tests at once, with both the order of tests randomised, and the order of words within each test.

Multivariate tests were then carried out in order to test whether the scores of the respective tests were statistically significantly different, with the results shown in the Table 10.

Table 10: Aurallex Pilot: Multivariate tests to establish similarities between tests

Multivariate Tests ^a							
Effect	Value	Hypothesis			Sig.	Partial Eta Squared	
		F	df	Error df			
tests	Pillai's Trace	.31	6.51 ^b	5	73.00	.000	.31
	Wilks' Lambda	.69	6.51 ^b	5	73.00	.000	.31
	Hotelling's Trace	.45	6.51 ^b	5	73.00	.000	.31
	Roy's Largest Root	.45	6.51 ^b	5	73.00	.000	.31

a. Design: Intercept
Within Subjects Design: tests

b. Exact statistic

They showed the value of Wilks' Lambda is .691, with a significance value of .000. Since the alpha level is below .05, the tests still had significant differences between each other.

In order to find where the significant differences lay, a series of pairwise comparisons were run. A process of elimination was carried out through these pairwise comparisons, by which the test with the most significant differences with the other tests would be removed, and then the repeated measures ANOVA would be re-run. This continued until four tests remained for which the pairwise comparison test suggested no significant differences. These were tests 3, 5 6 and 7, for which descriptive statistics are given in Table 11. The N=82 reflects students from all year groups combined.

Table 11: AuralLex Pilot: Descriptive statistics for remaining four Aurallex tests

Descriptive Statistics			
	Mean	Std. Deviation	N
Test 3	597.56	320.29	82
Test 5	571.34	313.78	82
Test 6	563.41	416.86	82
Test 7	510.37	340.64	82

The test of within-subjects effects (Table 12) confirms that these data do not violate the assumption of sphericity and the mean scores for vocabulary tests 3, 5, 6 and 7 were not statistically significantly different ($F(3, 243) = 1.906, p = 0.129$). Post hoc tests using the Bonferroni correction revealed that there were slight differences in the test scores, but that these were not statistically significant.

In order to understand better where any subtle differences between tests might have occurred, the pairwise comparisons (Table 13) sheds some light:

Table 12: Aurallex Pilot: Testing the assumption of sphericity for Aurallex 3, 5, 6 and 7

Tests of Within-Subjects Effects							
Measure: MEASURE_1							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
test	Sphericity Assumed	329054.88	3	109684.96	1.91	.129	.023
	Greenhouse-Geisser	329054.878	2.548	129126.784	1.906	.139	.023
	Huynh-Feldt	329054.878	2.638	124733.041	1.906	.137	.023
	Lower-bound	329054.878	1.000	329054.878	1.906	.171	.023
Error(test)	Sphericity Assumed	13980945.122	243	57534.754			
	Greenhouse-Geisser	13980945.122	206.413	67732.875			
	Huynh-Feldt	13980945.122	213.684	65428.157			
	Lower-bound	13980945.122	81.000	172604.261			

Table 13: Aurallex pilot: Pairwise comparisons - differences between tests

This pairwise comparison chart suggests that the similarities between tests 3, 5 and 6 are the closest, and Test 7 is slightly less close, but still with no statistically significant differences between scores.

Pairwise Comparisons

(I) test	(J) test	Mean Difference		Sig. ^a	95% Confidence Interval for Difference ^a	
		(I-J)	Std. Error		Lower Bound	Upper Bound
3	5	26.220	30.481	1.000	-56.220	108.659
	6	34.146	39.749	1.000	-73.359	141.652
	7	87.195	35.465	.096	-8.723	183.114
5	3	-26.220	30.481	1.000	-108.659	56.220
	6	7.927	44.355	1.000	-112.036	127.890
	7	60.976	30.823	.308	-22.387	144.338
6	3	-34.146	39.749	1.000	-141.652	73.359
	5	-7.927	44.355	1.000	-127.890	112.036
	7	53.049	41.659	1.000	-59.621	165.718
7	3	-87.195	35.465	.096	-183.114	8.723
	5	-60.976	30.823	.308	-144.338	22.387
	6	-53.049	41.659	1.000	-165.718	59.621

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Therefore it was concluded that the most reliable way to proceed in the main study was to simply rotate the four tests 3 5 6 7.

Testing for ceiling and floor effects

It was also necessary to test whether these tests were subtle enough to show progress throughout the period of the study – in other words, whether significant differences existed between the results for years 7, 8, 9 and 10. Although the present study only plans to track progress up until the end of year 9, it was felt that testing up to the end of year 10 should allow for any exceptional progress that might be shown by students by the time they were in year 9, and thus ensure that the tests did not have a ceiling effect for any participant towards the end of the study.

To this end, the sum of the scores, for tests 3 5 6 7 were compared with a one-way ANOVA. The descriptive statistics are given in Table 14:

Table 14: Aurallex Pilot: Descriptive statistics to investigate floor and ceiling effects

Descriptives								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
7	19	942.11	722.70	165.80	593.78	1290.43	50.00	2100.00
8	28	1428.57	809.48	152.98	1114.69	1742.46	.00	2950.00
9	26	1844.23	808.12	158.49	1517.82	2170.64	300.00	3400.00
10	28	3342.86	930.82	175.91	2981.92	3703.79	1300.00	5450.00
Total	101	1974.75	1217.70	121.166	1734.36	2215.14	.00	5450.00

The between-groups ANOVA score for the sums for tests 3 5 6 7 showed a significance of $p < 0.01$, (see Table 15) from which it can be concluded that there is a significant difference between the scores for each year group.

Table 15: Aurallex Pilot: difference between year groups

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	81464454.121	3	27154818.040	39.423	.000
Within Groups	66813664.691	97	688800.667		
Total	148278118.812	100			

As a result, the decision was taken that these four vocabulary tests are appropriate instruments to use and that by rotating them and comparing participants' tests scores on one test against another, it will be reasonable to draw conclusions as to the vocabulary progress made over the data collection period. After piloting, four were found to be of comparable difficulty and were selected for use in the present study.

Although my main listening proficiency tests increased in difficulty in line with my perception of the proficiency of the learners, I was able to avoid this challenge with the AuralLex and the GJT by creating a very wide range of accessibility suitable for all

English learners of secondary school French, measuring a vocabulary size of up to 2000 words (Milton, 2006, 2008), and encompassing the 20 key grammar points within the year 7 to year 11 curriculum (AQA, 2018). The final list of my AuralLex tests can be found in appendix 15.4.

3.4.7. Measuring grammatical knowledge

The grammaticality judgement test (GJT) is widely used in second language acquisition research and has been argued to be reliable (Gass, 1994; Shiu, Yalçın, & Spada, 2018). However, these tend to exist in a written modality. Although it is recognised that the aural modality appears to be harder for learners than the written one (Murphy, 1997), I wanted to use an aural approach in order to be consistent with the construct of listening proficiency to avoid the written form acting as a confound.

Therefore, I created ten sets of 20 sentences, each of which tapped into a specific grammar point from the French curriculum followed by the participants. In each set, 10 sentences were grammatically correct, and 10 sentences contained a single grammatical error (for example an error in word order or conjugation). As with the vocabulary instrument, my dual intentions were to create as many tests as possible of equal difficulty, and that they would also be able to show progress.

3.4.8. Piloting the AuraGram

As with the piloting of the vocabulary, the piloting of the grammar tests had four aims:

1. To ascertain whether the test made sense to participants and was clear and usable.
2. To ascertain whether scores from the test could be meaningful.
3. To create a range of grammaticality judgement tests of equal difficulty.
4. To ensure that the tests would indicate progress by presenting appropriate challenge throughout the testing period, without having a floor or ceiling effect.

Initial piloting

In the first GJT, 25 sentences were created, of which 14 were ungrammatical and 11 were grammatical. Grammatical structures were taken from the required grammar for the French curriculum followed by the participating schools, and ungrammatical sentences were formed according to common mistakes made by learners, as observed by the researcher. The participants would hear a sentence and then chose one of three options: 'Grammatical', or 'Ungrammatical', or 'I don't know.' One point was awarded for each correct answer, and no points for an 'I don't know'.

The test was piloted on year seven in May 2018 (i.e. after approximately eight months' tuition), as part of the first full pilot test, and the first aim was met, in that the grammar element of the full test presented no logistical problems to the participants. However, when the results were viewed alongside another element of the full pilot, that is, the test of listening and transcribing in French, questions arose. The transcriptions provided by many of the participants showed very little grasp of grammar, (e.g. 'j'aime mon cheval il je fais de la natation', where 'et' (and) was transcribed as 'il' (he)). Given the time-consuming nature of the piloting, at this juncture it was decided not to continue to pilot this particular GJT on other years.

A second GJT was compiled, but this was rejected after recording. It became clear that in spoken language, a judgement of grammaticality can become hazy depending on the prosody of the delivery, particularly when the insertion of pauses within the sentences almost suggested a grammatical error and then a correction of it. For example, the sentence 'je j'aime le foot' (I I like football [sic]), conceived due to an observation that many beginner learners fail to see j'aime as a chunk, and therefore add a further 'je' to the sentence, can be recited without a pause, and then be correctly judged as

ungrammatical, or can be said as ‘je... J’aime le foot’, where the pause brings with it an inference of a false start and a second attempt.

With both these conundrums in mind, questions were asked as to the validity of including a GJT within the model at all, and this was discussed at some length during the Transfer of Status viva. However, both examiners felt that it was important to include grammatical knowledge, not least because it appears on all theoretical models of listening (Anderson, 1990; Cutler & Clifton, 1999; Field, 2013) – often under the guise of ‘parsing’. Therefore it was decided to persevere with the inclusion of grammar in the model, and a third GJT was conceived, named ‘auragram’, to mirror the ‘aurallex’.

Conceiving the AuraGram

Given that the successful AuralLex had taken vocabulary from the GCSE curriculum, ten tests were devised, ideally of equal difficulty also based on the GCSE curriculum.

Twenty key grammatical items were taken from the AQA GCSE specification (AQA, 2016) and for each grammatical item, ten sentences were written, five of which were correct, and five of which contained mistakes which are commonly seen in the classroom of English learners of French. Care was taken to make each sentence as simple as possible, containing lexical items covered in the first half of the year seven curriculum, and many cognates, with the aim that lack of vocabulary knowledge would not confound the results. From this collection of sentences, 10 tests were compiled, each containing 20 sentences (one for each grammatical item), 10 of which were grammatical and 10 of which were ungrammatical. The sentences were then checked by two native-speaker teachers of French and a full consensus was reached on the appropriateness of the sentences.

Pilot participants

The new academic year had just begun by this stage, which meant that for practical reasons, the tests – with a total of 200 items – were piloted on 107 pupils in years 8, 9, 10, and 11 (who had between one and four years of French tuition respectively). Whole classes were tested simultaneously online during school language lessons, and up to an hour was allowed for testing. It was not feasible to pilot the tests on the new year 7 students as they were to be the participants of the main study. In order to remove any possible confound of test fatigue, the order of test delivery was randomised, and the order of sentences within each test was also randomised. Not all the year 9 participants completed all the tests, as a surprise fire drill interrupted testing in the year 9 group.

AuraGram pilot statistics

Tests were named A to J. To discover where the differences between the tests lay, a series of Wilcoxon signed rank tests were run. The initial findings are set out in Table 16, with the Wilcoxon signed rank scores reported using the Z statistic, and the asymptotic two-tailed p value below it, with non-significant results highlighted in green: values greater than 0.05 suggesting there is no significant difference between the outcome of the pair of tests.

These findings led to the removal of tests B, D and H: the Wilcoxon signed rank tests then showed no significant differences between the remaining tests (Table 17).

Table 16: Auragram Pilot: location of differences between Auragram tests

	A	B	C	D	E	F	G	H	I	J	
A Z			-4.909 ^b	-0.086 ^b	-4.195 ^b	-0.878 ^b	-1.411 ^b	-0.284 ^b	-3.752 ^b	-1.640 ^b	-1.272 ^b
Asymp sig			0.000	0.931	0.000	0.380	0.158	0.776	0.000	0.101	0.203
B Z				-4.215 ^c	-1.056 ^c	-4.235 ^c	-4.061 ^c	-4.993 ^c	-1.063 ^c	-3.768 ^c	-3.868 ^c
Asymp sig				0.000	0.291	0.000	0.000	0.000	0.288	0.000	0.000
C Z					-3.604 ^b	-1.137 ^b	-0.952 ^b	-0.035 ^b	-2.918 ^b	-1.354 ^b	-0.839 ^b
Asymp sig					0.000	0.256	0.341	0.972	0.004	0.176	0.402
D Z						-3.062 ^c	-3.340 ^c	-3.927 ^c	-0.411 ^c	-3.215 ^c	-3.192 ^c
Asymp sig						0.002	0.001	0.000	0.681	0.001	0.001
E Z							-0.113 ^c	-1.380 ^c	-3.057 ^b	-0.460 ^b	-0.485 ^b
Asymp sig							0.910	0.168	0.002	0.645	0.628
F Z								-1.032 ^c	-3.030 ^b	-0.338 ^b	-0.223 ^c
Asymp sig								0.302	0.002	0.735	0.824
G Z									-3.519 ^b	-1.168 ^b	-1.302 ^b
Asymp sig									0.000	0.243	0.193
H Z										-2.605 ^c	-2.822 ^c
Asymp sig										0.009	0.005
I Z											-0.583 ^c
Asymp sig											0.560
J											

Table 17: no remaining differences between Auragram tests

	A	C	E	F	G	I	J	
A Z			-0.086 ^b	-0.878 ^b	-1.411 ^b	-0.284 ^b	-1.640 ^b	-1.272 ^b
Asymp sig			0.931	0.380	0.158	0.776	0.101	0.203
C Z				-1.137 ^b	-0.952 ^b	-0.035 ^b	-1.354 ^b	-0.839 ^b
Asymp sig				0.256	0.341	0.972	0.176	0.402
E Z					-0.113 ^c	-1.380 ^c	-0.460 ^b	-0.485 ^b
Asymp sig					0.910	0.168	0.645	0.628
F Z						-1.032 ^c	-0.338 ^b	-0.223 ^c
Asymp sig						0.302	0.735	0.824
G Z							-1.168 ^b	-1.302 ^b
Asymp sig							0.243	0.193
I Z								-0.583 ^c
Asymp sig								0.560
J								

Floor and ceiling effects

It was also necessary to ascertain whether these tests were subtle enough to show progress throughout the period of the study – in other words, whether the tests had a floor or ceiling effect for all participants, from the least experienced and least able learners, to the most experienced and most able learners. Descriptive statistics demonstrate (Table 18) that by the start of year 8, no participant scored less than 1/20, and that the score of 20/20 was only achieved in two tests, and only by participants in year 11.

The descriptive statistics suggest that even during the first year of the study, the auragram test will be sensitive enough to demonstrate some grammatical knowledge among nearly all participants.

Table 18: Descriptive statistics for Auragram tests, by year group

Year Group		Test A	Test C	Test E	Test F	Test G	Test I	Test J
year 8	N	21	22	20	22	22	22	21
	Mean	7.86	6.27	8.40	8.18	8.14	8.00	7.86
	Std. Deviation	2.85	2.47	2.23	1.68	2.23	2.53	2.73
	Minimum	1	3	5	6	5	3	4
	Maximum	12	12	12	11	13	12	14
	Range	11	9	7	5	8	9	10
year 9	N	25	24	24	26	24	26	26
	Mean	5.84	7.08	7.88	7.77	7.42	7.65	7.38
	Std. Deviation	2.72	3.24	2.27	2.14	2.47	2.91	2.59
	Minimum	1	2	4	4	3	3	2
	Maximum	11	13	12	12	13	14	12
	Range	10	11	8	8	10	11	10
year 10	N	27	27	26	27	27	27	27
	Mean	9.26	10.26	9.65	9.63	9.44	10.22	10.74
	Std. Deviation	2.98	3.32	2.87	2.50	3.00	2.76	3.06
	Minimum	3	3	3	5	5	6	3
	Maximum	15	18	17	13	17	16	16
	Range	12	15	14	8	12	10	13
year 11	N	28	29	29	30	30	28	29
	Mean	11.79	11.38	11.17	10.83	10.37	10.86	10.45
	Std. Deviation	3.88	4.20	3.77	3.39	3.09	3.67	3.93
	Minimum	3	5	5	4	2	5	4
	Maximum	18	20	20	19	15	18	18
	Range	15	15	15	15	13	13	14

External validity

In order to assess the Auragram's external validity, the pilot participants also took the French version of the Oxford Placement Test (OPT) (Allan, 1995), a widely validated test of French grammatical knowledge, with the aim of examining whether the outcomes of the two tests were related. Given the non-parametric distribution of the data produced by the OPT scores, a Spearman's rank-order correlation was run to determine the relationship between the participants' Auragram scores, and those of the OPT. There was a strong positive correlation between the two sets of scores, which was statistically significant ($r_s = .687$, $p = .000$).

Pilots therefore indicated that six of these tests were of equal difficulty, and that the tool was able to differentiate between learners in years 7, 8, 9 and 10. For full details of piloting, see appendix **Error! Reference source not found.**

3.4.9. Connection between vocabulary and grammar

It is important at this juncture to acknowledge the logical connection between vocabulary and grammatical knowledge (Bulté & Housen, 2020). Grammatical knowledge cannot be tested without some level of vocabulary knowledge, and as such they cannot be operationalised as entirely independent of each other. As much as possible, I incorporated cognates as the lexical items in my GJT, but nonetheless, when looking for supporting or competing relationships between these two variables, such interdependencies should be kept in mind. The final list of AuraGram tests can be found in appendix 15.5.

3.4.10. Measuring strategy use

I piloted the Metacognitive Awareness Listening Questionnaire (MALQ) (Vandergrift et al., 2006) (appendix 15.1.1), which is a well-established instrument with questions which speak to a range of socio-cognitive processes involved in listening proficiency: these are

Problem-Solving, Mental Translation, Person Knowledge, and Directed Attention. In response to students' confusion, I removed the double negative from one question (and correspondingly reversed the scoring mechanism for that question); otherwise, I felt able to use the instrument as is.

3.5. Data collection procedure

At each of the 20 time points, tests assessed listening proficiency via a 'paused translation' protocol), vocabulary and grammar. In addition, participants were asked to report on their strategy use, and their self-efficacy and motivation, and the four case study participants were also interviewed, which included stimulated recall. This is set out in Table 19.

Table 19: Data collection procedure

	3x paused translation tests	AuralLex	Aural GJT	MALQ	Semi-structured interview	Stimulated recall
76 standard participants	✓	✓	✓	✓		
4 case study participants	✓	✓	✓	✓	✓	✓

This section will address how I used my instruments to generate data.

3.5.1. Listening proficiency

At each time point, three different passages from three difficulty levels were played to the participants (for how I determined difficulty levels, refer back to section 3.4.1, above).

Three tests were chosen to avoid floor and ceiling effects as much as possible; to reduce text effects such as pre-existing knowledge, and negative effect of initial misinterpretation of gist (Yeldham, 2017b); and so that at each time point two could be repeated from a previous time as anchor points, for the purposes of Rasch modelling (see section 3.3.2).

In Table 20, which sets out the testing schedule, repeated tests are in bold. At least one text at each time point was on the theme of the students' current 'main topic' – that is to say, the theme that the participants had been studying at the time of the test.

Table 20: Testing protocol

Time	Yr	Date	Test 1	Test 2	Test 3	Main topic
1	7	Oct-18	AS7	AS4	AS8	me
2	7	Dec-19	AS8	St23B	F25	school
3	7	Feb-19	G78	St23B	E111	hobbies
4	7	Mar-19	F9	AS7	E26	sports
5	7	May-19	F28	F25	E3	town
6	7	Jul-19	G32	St23B	E56	holidays
7	8	Sep-19	F46	F9	E26	hobbies
8	8	Oct-19	St212A	E3	D8	hobbies
9	8	Dec-19	F25	E65	D25	Paris
10	8	Feb-20	E111	F9	D3	identity
11	8	Mar-20	E3	D47	C48	my home
12	8	May-20	E26	D8	C17	events
13	8	Jul-20	E56	D25	C55	celebs
14	9	Sep-20	E111	D31	C48	social life
15	9	Oct-20	E26	D3	C2	social life
16	9	Dec-20	E3	D8	C41	healthy living
17	9	Feb-21	D47	C48	C11	jobs / future plans
18	9	Mar-21	D25	E111	St3AP	holidays
19	9	May-21	E56	D3	St351A	holidays
20	9	Jul-21	C48	St342A	D8	ethics / the world

For each test, the participant read screens such as the ones in

Figure 8, which gave a schema for the respective passages.

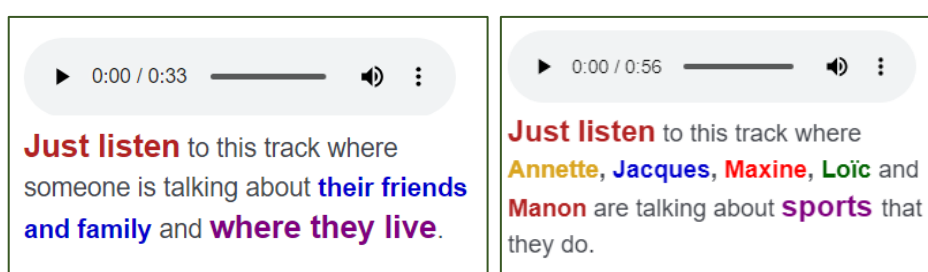


Figure 8: Establishing the schema in the listening passage

They then listened to the passage in its entirety. The participant then listened again in short chunks of up to 10 words, and translated what they heard into English.

3.5.2. Vocabulary knowledge

After participants had completed the three listening tasks, they were given the AuralLex vocabulary test. There were four tests of similar difficulty (see section 3.4.5 for how these were created), which rotated (in other words, test 1 at Time 1, test 2 at Time 2, test 3 at Time 3, test 4 at Time 4, then back to test 1 at Time 5, test 2 at Time 6, and so on). This is illustrated in Table 21.

Table 21: Which AuralLex test when?

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AuralLex test #	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

After an introductory screen (Figure 9), participants listened to the 48 words one by one, the order randomised for each participant by the software (Figure 10). After playing each word, participants had to respond either ‘I know or can use this word’ or ‘I do not know this word’; there was no possibility just to skip without responding.

This section is about your knowledge of French vocabulary. It should take less than five minutes to complete.

Each time, you will hear a word. Some of these words are real French words. Some are invented by are made to sound like real French words. Each time, say whether it is a word you know or can use.

For example, in English, if you heard the word 'football', you would choose 'I know or can use this word', but if you heard the word 'sloskon', you would chose 'I do not know this word'.

You can listen to each word a maximum of two times.

Figure 9: AuralLex introductory screen

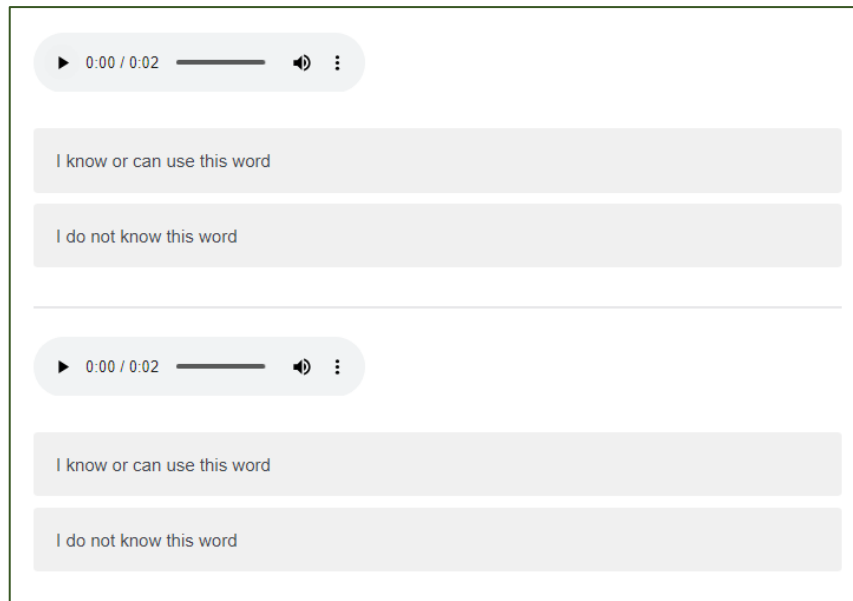


Figure 10: Part of main AuralLex test screen

Scoring mirrored that of the X-Lex (Milton, 2006): fifty points were awarded for each time a participant claimed ‘I know or can use this word’ if it was a real word, and 250 points were deducted if a participants claimed to recognise or be able to use a non-word (i.e. a false positive), in order to correct for guesswork. From this an estimate of the participants’ vocabulary size (out of the two thousand most frequent French words) was made.

3.5.3. Grammatical knowledge

Having completed the AuralLex, in the AuraGram element of the test battery, participants took an aural grammaticality judgement test. After a preamble (

Figure 11), they were played 20 sentences in French and, after each one, were asked to indicate whether they believed the sentence was grammatical or not grammatical (

Figure 12). As with the AuralLex, the order of the sentences was randomised so that each participant would hear the sentences in a different order. Each participant was awarded a

score out of 20, according to how many sentences they had correctly judged grammatical or ungrammatical.

Each time, you will hear a sentence. Some of these words are grammatical French sentences. Some have a mistake in them. Each time, say whether the sentence is grammatical, or it is ungrammatical, or you don't know.

For example, in English, the sentence 'I go to the park' is **grammatical**, and the sentence 'I to the park go' is **ungrammatical**. 'He is playing football' is **grammatical**, and 'he are plays football' is **ungrammatical**.

Don't overthink it! Try to give a very quick response! You should listen to each sentence a maximum of two times.

Figure 11: AuraGram preamble

▶ 0:00 / 0:03 Listen to this sentence. Is it...

Grammatical

Ungrammatical

I don't know

▶ 0:00 / 0:03 Listen to this sentence. Is it...

Grammatical

Ungrammatical

I don't know

Figure 12: AuraGram main test

3.5.4. Metacognition

The MALQ (Vandergrift et al., 2006) came immediately after the listening exercises, as recommended by Cross & Vandergrift (2015), to link it to a specific listening experience. The preamble is given in Figure 13 and an example of the main questionnaire in Figure 14. I collated scores from a Likert scale of 1-6 to create scores for each participant on the elements of Problem-Solving, Directed Attention, Person Knowledge and Mental Translation.

This section is to find out about how you go about listening tasks in French.

Give a rating for each of the statements. There are 21 statements. This section should take about two minutes to complete.

Figure 13: Preamble to the Metacognitive Listening Awareness Questionnaire

I translate word by word, as I listen.

Strongly agree Agree quite a lot Agree a bit Disagree a bit Disagree quite a lot Strongly disagree

I translate in my head as I listen.

Strongly agree Agree quite a lot Agree a bit Disagree a bit Disagree quite a lot Strongly disagree

As I listen, I quickly adjust my interpretation if I realise that it is not correct.

Strongly agree Agree quite a lot Agree a bit Disagree a bit Disagree quite a lot Strongly disagree

Figure 14: Sample of main part of Metacognitive Listening Awareness Questionnaire

3.5.5. Semi-structured interviews

The semi-structured interviews were held as soon as possible after each test battery was held (usually the same day) and lasted around half an hour each. A simple question protocol (appendix 15.9) was formulated, with additional, unscripted questions posed to the interviewee where the researcher deemed them appropriate. During the interview, I always carried out a stimulated recall of one of the paused translation passages, whereby I called up what they had written, and explored with them the reasons for their answer. I always used one of the unrepeated tests (those which are not in bold type in Table 20); the listening tests which had been designated as anchor points were avoided for stimulated recall in case detailed discussion of the track improved the participants' performance on future occasions.

3.6. Data analysis procedure

A deductive approach to data analysis begins with theory, sets up hypotheses and gathers evidence in order to falsify – or not – these hypotheses (Jordan, 2004). CDST takes a more inductive approach, whereby researchers attempt to categorise phenomena and look for systematicity only after data collection has been completed (Dörnyei, 2014; Hiver & Al-Hoorie, 2020). The upshot of this is that, when examining a system that is complex, dynamic and iterative, where patterns are unpredictable, it is necessary to take a broad, holistic approach. This presents a methodological challenge in that a lot of data are collected, some of which turns out to be superfluous.

3.6.1. RQ1: listening error analysis

Research question 1 asked, what do listening errors of school-aged students of French reveal about a) their listening processes, and b) the development of their listening proficiency over a three-year period? A diagram setting out my approach to data analysis for RQ1 is available in Figure 15. Definitions of terms can be found in the codebook summary, appendix 15.10.

As previously discussed in section 2.5.2, to the best of my knowledge, when I designed the research, Dong (2016) and Chang (2017) were the only researchers who had applied CDST theory to L2 listening, and these researchers used the scores from established tests to create data. They do not address error analysis. The literature review also explores papers which analyse errors in listening through paused transcription (not translation), and I referred to these closely for potential data analysis models, with varying degrees of success. Neither Wong, Leung, Tsui, Dealey, and Cheung (2021) nor Cross (2009) gave much indication of their data analysis method, simply stating ‘the data were analysed’ (Wong et al 2021:4) and that the researchers ‘identified errors and allocated them to

codes' (Cross 2009:38). The methodology of both Leonard (2019) and Carney (2021) revolves around the use of idea units which I had previously ruled out, given that the reductionist approach hinders analysis from a CDST perspective, coupled with its divergence from Field's model (2008b, 2013). Better were both Lange and Matthews (2023) and Sheppard and Butler (2017), who explicitly acknowledged a qualitative, recursive approach to error analysis, although both also state that they began with the quantitative score, and only qualitatively analysed the lowest scoring transcriptions.



Figure 15: Data analysis process for Research Question 1

In CDST research into L2 writing, however, both Larsen-Freeman's (2006) and Lowie and Verspoor (2018) analyse the entire corpus of data which is produced by their participants. As such, I combined the qualitative and recursive listening error analysis of Lange and Matthews (2023) and Sheppard and Butler (2017) with the thoroughness of Larsen-Freeman (2006) and Lowie and Verspoor (2018), to analyse the entire body of the

paused translation data (consisting of around 400,000 words), taking a reflexive thematic analysis approach (Braun & Clarke, 2020; Byrne, 2021). My reflexive diary kept throughout the data analysis process is available in appendix 15.6 and a collection of screenshots of the coding process at various iterations is available in appendix 15.11. The coding began deductively with three broad categories of segmentation, vocabulary and grammar, after Field's model (2008b, 2013), and all sub-codes were developed inductively and recursively as I made my way through the dataset. Codes were checked with an interrater (a fellow teacher of French, also qualified to Master's level in applied linguistics), who independently coded 10% of the data: initially there was 92% agreement; disagreements were resolved through discussion. Many of the entries provided were double-coded, but each code was allocated to an entire track and never to an individual word. The coding work allowed me to generate the themes which are set out and analysed in the accompanying findings chapter.

3.6.2. RQ2: relationships between subsystems

Research question 2 asks: How do the relationships between listening proficiency and the subsystems of language knowledge (vocabulary knowledge and grammatical knowledge) and metacognition (Person Knowledge, Problem-Solving, Mental Translation and Directed Attention) develop over three academic years? To address this question, I followed the procedure set out in Figure 16.

Sample mark schemes for the paused translation are given in appendix 15.7. Once marked, the scores were converted to percentages, and then, by means of a Rasch model (section 3.3.2 (Verspoor, Lowie, & Van Dijk, 2008)), I converted the paused translation scores onto a common scale (see appendix 15.8). This resulted in a single score out of

100 for each time point for each participant. The remainder of the tests did not necessitate Rasch modelling as the scales remained constant throughout the data collection.

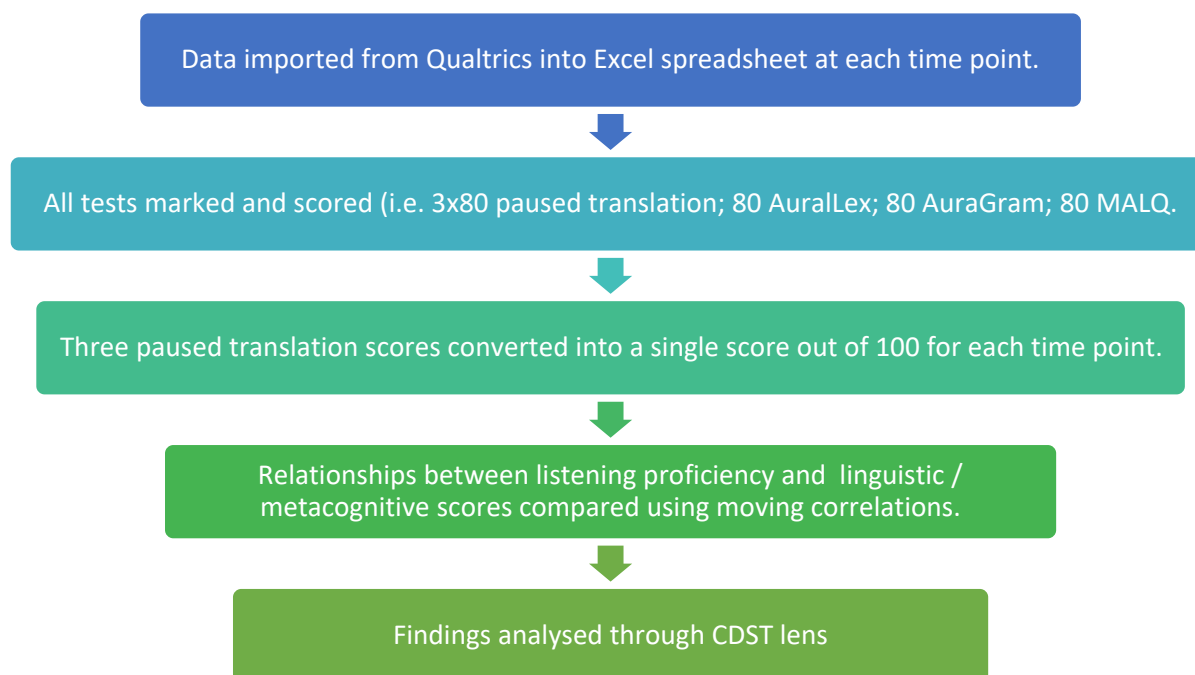


Figure 16: Data analysis procedure for Research Question 2

To start with, the nature of the data within each individual subsystem was explored. According to CDST, variability can be a meaningful source of information on the *process* of language development (Yu & Lowie, 2020) and points to a dynamism of development (Van Geert & Van Dijk 2002). The large numbers of data points not only allow us to examine the extent to which progress is linear, or not, it also allows us to interrogate both the nature and the implications of any non-linearity. For example, when we identify changes in the nature of a developmental trajectory – such as increased fluctuations in scores shortly before a jump in proficiency (Smit, Dijk, de Bot, & Lowie, 2022), or a change in the underlying subsystems, we are able to deepen our understanding of the processes of learning under investigation.

Mean scores for each subsystem, overlaid with a trendline (polynomial to a factor of 6), were reported. While acknowledging CDST's emphasis on variability, the purpose of the

trendline is to simplify the pattern to such an extent to allow the capture of general tendencies (van Dijk, Verspoor, & Lowie, 2011).

Individual subsystems: moving min-max graphs

Further patterns of variability become apparent by means of the moving min-max graph (Van Geert and van Dijk, 2002), which can trace changes and degrees of variability during development. Van Dijk et al. (2011) recommend setting a window of five time-points, which offers hints at potential developmental jumps and slumps. From the moving min-max graphs, our understanding of the non-linear progress is deepened through analysis of the ‘bandwidth’ (the difference between the minimum and the maximum mean score within a window of five time points). In other words, the difference between the minimum score and the maximum score in some windows is wider than in others, demonstrating a wider range of mean proficiency at some points. Once the initial graph has been produced, analysis involved exploration of different phases of learning to probe the underlying principles and mechanisms of change. Allocations of phases can be argued to be subjective, and therefore all graphs were discussed with two interraters (one of my doctoral supervisors, and a fellow languages teacher also qualified to Master’s level in applied linguistics) and differences of opinion were resolved through discussion.

Relationships between subsystems: moving correlations

In order to explore the changing relationships between the different subsystems over time, I used a moving correlation approach (Verspoor & van Dijk, 2011, Lowie & Verspoor, 2018). This is a CDST descriptive statistics technique which aims to investigate whether relationships are ‘supportive’ (subsystems grow simultaneously because they support each other), ‘competitive’ (one subsystem develops and another withers), or ‘conditional’ (one can only grow after another has developed). I followed the

instructions set out by Verspoor, Lowie, van Geert, van Dijk, & Schmid (2011): all numbers were normalised – recalculated on a standard scale from 0 to 1 – in order to be able to represent them graphically on a shared y-axis. Correlations were calculated using both Microsoft Excel and SPSS, and I adhered to the standard measures whereby $\pm .1$ represents a small effect, $\pm .3$ represents a medium effect and $\pm .5$ represents a large effect (Lowie, 2012).

The aim of the moving correlation methodology is twofold. Firstly, it allowed me to coordinate changes in the relationships under examination, and crucially how the trends in those relationships might change over time. Secondly, it renders the interactions between the subsystems under examination more visible. This direction of focus towards change over time deviates from a more traditional focus on the value of correlation in a population; the upshot of which is that although the work might be construed to be quantitative, it is descriptive only, and MacIntyre and Ducker (2022) and Lowie (2023) counsel against significance tests, stating they are not appropriate in this context.

Moving windows of correlations were created using windows of three data points.

Verspoor et al. (2011) recommend beginning with a window of five data points; longer windows are more likely to produce stable correlations than shorter ones. When deciding on window size, a deciding factor is to find a dynamic analysis that is meaningful for the variables and the participants of the study (Bulté & Housen, 2020). It is important to understand the specific events particular to them which might influence the changing relationships between the variables (MacIntyre & Ducker, 2022). I chose a window of three which equated to one academic term, and reflected how much can change for a young person over one third of an academic year, as well as reflecting the length of the first covid lockdown. Although in moving correlation analysis this is relatively short, other researchers have used windows of three (e.g. Fogal & Verspoor, 2020).

3.6.3. RQ3: case studies

Research question 3 asks: To what extent, and how, do the answers to questions 1 (errors and what they reveal about the listening process) and 2 (relationships between subsystems) vary between individuals, and how do individual patterns compare to those seen at group level? Are there other factors in the students' experience which are relevant to their listening comprehension development? This allowed me to address the CDST assertion that there is no such thing as a 'typical' participant (Lowie & Verspoor, 2018; Molenaar & Campbell, 2009), and to explore the experiences of four case studies, without falling into the trap of generalising from group to individual (MacIntyre & Ducker, 2022). This challenge is known as the ergodicity problem.

In order to answer these questions, in addition to completing the full test battery at each time point, eight participants initially participated in interviews immediately after each test battery. One left the school part way through the data collection period. To address RQ3, I began by examining the descriptive quantitative data explored for RQ2, but only the data generated by Helen, Theo, Fran and Ryan, four of my case studies. However, in addition to construction developmental trajectories for the individual subsystems and moving correlations for the relationships within the linguistic and metacognitive subsystems (see section 3.6.2 for methodology), I also used change point analysis software (Taylor, 2000). For practical reasons – for example, word count and risk of repetition – the other three case studies are not analysed in the present thesis. Initially I had two high prior attaining boys, two high prior attaining girls, two lower prior attaining girls and two lower prior attaining boys. I kept one case study in each category and my reasons for removal were as follows:

- Second high-prior attaining boy left the school

- Second high-prior attaining girl was remarkably uncommunicative at interview and as such provided very little qualitative data.
- The same was true of the second lower prior attaining boy.
- My remaining low prior-attaining girl was less uncommunicative than the others, but I made a judgement that Fran's highly reflexive insights into her own learning, coupled with the experience of her bilingualism provided a case study of great value.

Change Point Analysis

Dedicated change point analysis software (Taylor, 2000) is favoured by CDST researchers (e.g. Baba, 2020; Baba & Nitta, 2014; Han & Hiver, 2018; Hepford, 2020; MacIntyre & Ducker, 2022) to identify moments of significant change within chronological data. It uses a bootstrapping statistical approach with a confidence interval with 10,000 bootstraps without replacement. In other words, the software generates a series of randomised variations on the data in order to estimate the probability that a change has occurred. Once the analysis has taken place, the software creates a graph illustrating the change point and reports on the % confidence that the change was statistically significant. It is important to note, however, that such statistically inferences are not the same as critical incidents during the data collection period (Han & Hiver, 2018), which are more likely to be detected through qualitative methods.

Thematic Analysis of interview data

In addition to the quantitative data, to address RQ3 I made use of the transcripts of interviews held with the participants at each time point (for interview schedule see appendix 15.9), which always included a stimulated recall of one of the paused translation tasks. The qualitative analysis again followed a reflexive thematic analysis

approach (Braun & Clarke, 2020; Byrne, 2021). I triangulated the various data sources (min-max, moving correlations, change point analysis plus interview data) in order to build a picture of each individual's developmental trajectory and the changing relationships between the subsystems over time.

3.7. Ethical considerations

3.7.1. Minimising harm

To protect the identity and confidentiality of my research participants, I have not named the school or the city in which the research has taken place. All the full cohort participants are identified solely as a four-letter code and these codes were drawn up very early in the research, meaning that confidentiality was ensured for all colleagues and peers who might have looked at my data in any form. The case studies were all given pseudonyms.

Consent was gained twofold: the university department's ethics committee agreed that the main data collection was consistent with normal classroom activities and needed opt-out rather than opt-in parental consent. As such, letters were sent to the families of all children in year seven explaining the research project, and giving the opportunity to discuss it further with the researcher if they required, but they were also told that if they did not explicitly opt out, the data that students provided with regard to the ongoing listening comprehension tests would be used. An additional letter was sent home asking for volunteers to opt in for the case study element of the work, from which eight case study students were initially chosen (although one left the school half way through the data collection period). In all communications with families, it was emphasised that students could withdraw from the study at any time. All ethics approval documents are available in appendix 15.12.

3.7.2. Position as a teacher-researcher

As a member of staff in the school, I held a privileged position vis-à-vis the participants: I was aware that they might feel pressured into participating, or might tell me what they thought I wanted to hear. Piloting showed that some students were anxious about ‘passing’ or ‘doing well’ and I took care to reiterate to the participants that this was research and that I was interested simply in their performance, without judgement.

I also had a privileged relationship with other teachers, from whose students I was collecting data. I needed to be aware of ‘wearing two hats’ while conducting the present study, and made it clear that I would not share any data from the study until after the data collection period had passed.

Within my role as a teacher at the participating school, I had a safeguarding obligation towards all students, and to that end, the question in the self-efficacy questionnaire ‘is there anything that might have affected your performance today’ necessitated a follow-up comment stating that I cannot keep any potential disclosures confidential. There were three occasions over the data collection period where I followed up comments that were given in response to this question for welfare reasons; those comments will be kept confidential here as they are not relevant to the findings of the research.

3.7.3. Data and record-keeping

Data were collected through the university’s secure Qualtrics platform, to which only I had the password. Once downloaded in spreadsheet form into the school’s secure and approved file server, and when anonymised, transferred back onto my university Nexus 365 server as per university policy. Interview data in the form of audio files, and subsequently transcriptions, were also stored on my university Nexus 365 server as per university policy (Oxford, 2022). No data were kept in paper format.

3.8. Problems arising

3.8.1. The impact of the Covid-19 pandemic

Table 22 summarises events surrounding online learning due to the pandemic.

Table 22: Location of teaching and data setting throughout the data collection period

Time	Date	Location of data collection	Type of teaching in term preceding test	Interview data collection setting	Total completed	Total excluded	Total valid
T1	Sep-18	School	Face to face in school	one to one in school	103	0	103
T2	Dec-18	School	Face to face in school	one to one in school	98	4	94
T3	Feb-19	School	Face to face in school	one to one in school	101	4	97
T4	Mar-19	School	Face to face in school	one to one in school	105	3	102
T5	May-19	School	Face to face in school	one to one in school	96	6	90
T6	Jul-19	School	Face to face in school	one to one in school	98	6	92
T7	Sep-19	School	None - long school holidays	one to one in school	94	10	84
T8	Oct-19	School	Face to face in school	one to one in school	94	7	87
T9	Dec-19	School	Face to face in school	one to one in school	81	2	79
T10	Feb-20	School	Face to face in school	one to one in school	90	3	87
T11	Mar-20	Hybrid	Face to face in school	one to one in school, apart from Theo, via zoom	89	9	80
T12	May-20	Home	Remote asynchronous	phone	83	6	77
T13	Jul-20	Home	Remote asynchronous	phone	80	6	74
T14	Sep-20	Home	None - long school holidays	one to one in school	97	12	85
T15	Oct-20	Home	Face to face in school	one to one in school	98	13	85
T16	Dec-20	Home	Face to face in school	one to one in school	90	15	75
T17	Feb-21	Home	Remote – live + asynchronous	video call	90	12	78
T18	Mar-21	School	Mix – remote live, remote asynchronous then live in school	one to one in school	95	9	86
T19	May-21	School	Face to face in school, with small groups self-isolating	one to one in school	90	13	77
T20	Jul-21	School	Face to face in school, with small groups self-isolating	one to one in school, apart from Helen, via Google meet	92	12	80

In English schools, a period of online learning began in March 2020. In the participating school, this took the form of work being set to be completed in the students' own time; during the initial lockdown there were no live, online lessons. This affected data collection at time 11 (March 2020), Time 12, (May 2020) and Time 13 (July 2020).

However, I continued to make and set listening surveys, and most participants completed them. Of course, as these were done at home, I had no control over the environment in

which these surveys were taken, the device used, and only minimal control over the timing.

In September 2020 all students returned to school, so teaching returned to normal for Time 14 (September 2020), 15 (October 2020) and 16 (December 2020), but social distancing within school meant that computer rooms were not available, so students still had to do the listening surveys at home – therefore I still had no control over environment, device or timing.

In January 2021, a second lockdown began. Although this time there was live online teaching via Google Meet, the same data collection problems arose. This affected Time 17 (February 2021). Students returned to face-to-face school very shortly before Time 18 in March 2021, whose data collection took place in computer rooms in school, in the same fashion as Time 1 to Time 10. Time 19 took place as anticipated in school, with face-to-face teaching for all classes, and all teachers were present. Time 20 took place the day after a school trip which had resulted in a covid outbreak, so again data collection was compromised as some participants had been obliged to self-isolate.

From Time 13, I added two additional questions to the survey to ask participants to self-report on whether they had completed the survey in a single sitting, and a yes / no question on whether they had experienced any distractions; and if so, what these were.

These findings can be summarised in Table 23 as follows:

Table 23: Additional details on remote data collection

	Single Sitting	Several sittings	Distractions	Not distracted	Number who did in single sitting with no distractions
Time 13	61	12	41	32	29
Time 14	62	12	39	35	33
Time 15	59	18	27	50	39
Time 16	56	15	31	40	32
Time 17	63	9	27	45	41

The various distractions mentioned by participants paint a picture of typical domestic life, where pets, siblings, parents, and domestic chores all interfere with schoolwork, as well as technical issues such as problematic wifi and omnipresent mobile phone notifications.

3.8.2. Test fatigue and attrition

Bearing in mind that participants were completing the survey within lesson time, I had not anticipated much attrition, apart from those students who left the school or withdrew from the study during the three academic years of data collection. However, from Time 2 a small number of students effectively withdrew from the study by appearing to complete each survey, but actually leaving very many of the listening questions blank.

Another category of participants appeared to complete the survey, but, after scrutiny, it was clear that they did so by not listening to some or all of the listening tracks. By adding up the total time of all the listening tracks embedded in each survey, I could easily compute a minimum number of seconds needed by a participant in order to complete it satisfactorily. Timings data provided by Qualtrics suggested that at each time point, a number of participants had spent less time on the entire survey than the total duration of the listening tracks.

Therefore, two key categories for exclusion became clear from as early as Time 2: those who had not provided any listening answers, and those for whom I could conclude that they had not listened to all the tracks. Both these causes of attrition began pre-pandemic, but were exacerbated by doing the surveys at home. Full details of the justification for exclusions are given in appendix 15.13.

3.9. Summary of chapter

This chapter has set out the methodology of the present study. It initially addressed my own ontological position and then discussed the different ways in which the developmental trajectory of L2 listening could have been investigated. It then provided detail on my methodological decisions both before and during the project, and ended with a discussion on the challenges arising during the project. In conjunction with the Literature Review, it puts into context the findings the research questions, which will be addressed in the next six chapters.

4. Findings 1: errors and listening processes

Research question one asked, what do listening errors of school-aged students of French reveal about a) their listening processes, and b) the development of their listening proficiency over a three-year period? Recall that a diagram setting out my approach to data analysis for RQ1 is available in Figure 15.

4.1. Introduction to the chapter

This chapter will examine types of listening difficulties that have emerged from the paused translation data. Participants' paused translations of aural input of French will be analysed, according to Field (2008b)'s theoretical framework. Very broadly, we see two types of errors: those which are linked to sound of the input – these will be discussed first – and those linked to the content. The chapter will begin with a discussion of segmentation errors (sound), followed by vocabulary errors (sound, then content) and grammatical errors (mostly content). Throughout, three themes will be apparent:

- For many participants on many occasions, the sound of a word is the driver of attempts to build meaning, sometimes in spite of common sense.
- There is a complex and dynamic interaction between elements in each attempted translation as participants attempt to form meaning.
- There is a wide range of approaches taken by participants to compensate for incomplete understanding.

At the very outset it is important to reiterate, as discussed in the literature review, that the fluidity and dynamism inherent in the listening process mirrors that of the conceptual framework of the present study. Examples of one phenomenon might also belong in other

sections, and sections often inevitably overlap, and inform each other. I have teased out themes and examples to create a logical structure, but do not claim that the result is impeccably tidy.

4.1.1. Explanation of chapter structure

A key theme of this chapter is the acknowledgement of the interactions between the different types of error. Recall that the literature review discussed the small body of scholarly work in L2 listening error analysis. There does not appear to be consensus as to ideal categorisation of such errors, with each team of researchers building their own model, driven by a theoretical framework, yet adapted as a result of their own study. Uniquely, Wong, Leung, Tsui, Dealey, and Cheung (2021) base their work on that of Bond (1999), which does not distinguish between first and second language listening, and forms six categories: a) simple consonant error, b) complex consonant error, c) vowel error, d) misperception of the shapes of words, e) lexicon error, and f) syntax-related/syntactic error (p1024). Others, like the present study, invoke Field (2008b) in a variety of ways: Cross (2009) cites Field's broad categorisation of 'process problems' – i.e. those due to deficiencies in the skill of listening *per se* – and 'text problems', which concern themselves with inadequate knowledge of the L2, but also adds his own category of 'intrusion', whereby the listener's L1 interferes with the L2 decoding. Leonard (2019) asks whether decoding errors are phonetically similar to the input, semantically appropriate based on the context, both, or neither. Lange and Matthews (2023) concern themselves with L2 listeners' difficulties in decoding known lexis, and also create a phonological category (L1 phonological influence) and three linguistic ones (collocation unfamiliarity, syntactic knowledge, and difficulties using co-text). Carney (2021) also addresses L1 phonological influence, alongside English speech modifications and

misinterpretation of context. Sheppard and Butler (2017)‘s categories were ‘segmentation errors’, ‘phonemes’, ‘unknown words and phrases’, and ‘top-down fabrications’.

In all the studies cited above, there are cases where the error classification inevitably overlaps, and the present study is no different. From the tangle of data that has arisen, every effort has been made to attribute the errors according to Field’s model of listening (2008b, 2013) – that is, by attributing errors to categories of segmentation, vocabulary, and grammar. These categories also form the structure of this chapter. Indeed, as Broersma and Cutler (2008:23) state: *‘Word recognition is the central component of language processing; there is no sentence without the words comprising it, and phonemic contrast only occurs because it distinguishes words.’* From this perspective, it follows that a vocabulary category should be central, and all other sections, including phonology will flow out from it. This stands in contrast with many of the studies cited in the previous paragraph which tend to view phonology as a stand-alone section, distinct from the other categories. I have preferred to split up the vocabulary and grammar sections into phonologically driven errors, textually driven errors, or both, mirroring Field’s concepts of ‘process problems’ and ‘text problems’. The structure of the rest of the chapter, then, is set out in Table 24 and Figure 17.

Table 24: More detailed chapter structure breakdown

Phonologically driven	Segmentation errors	Whole track	Section 4.2
	Vocabulary errors	Individual word boundaries Liaison	Section 4.3
		Saliency of specific sounds Metathesis	
Textually driven		Lexical sets Transfer of training	Section 4.4
Both phonologically and linguistically driven	Grammatical errors	False friends Function and content words Function and content morphemes	Section 4.5

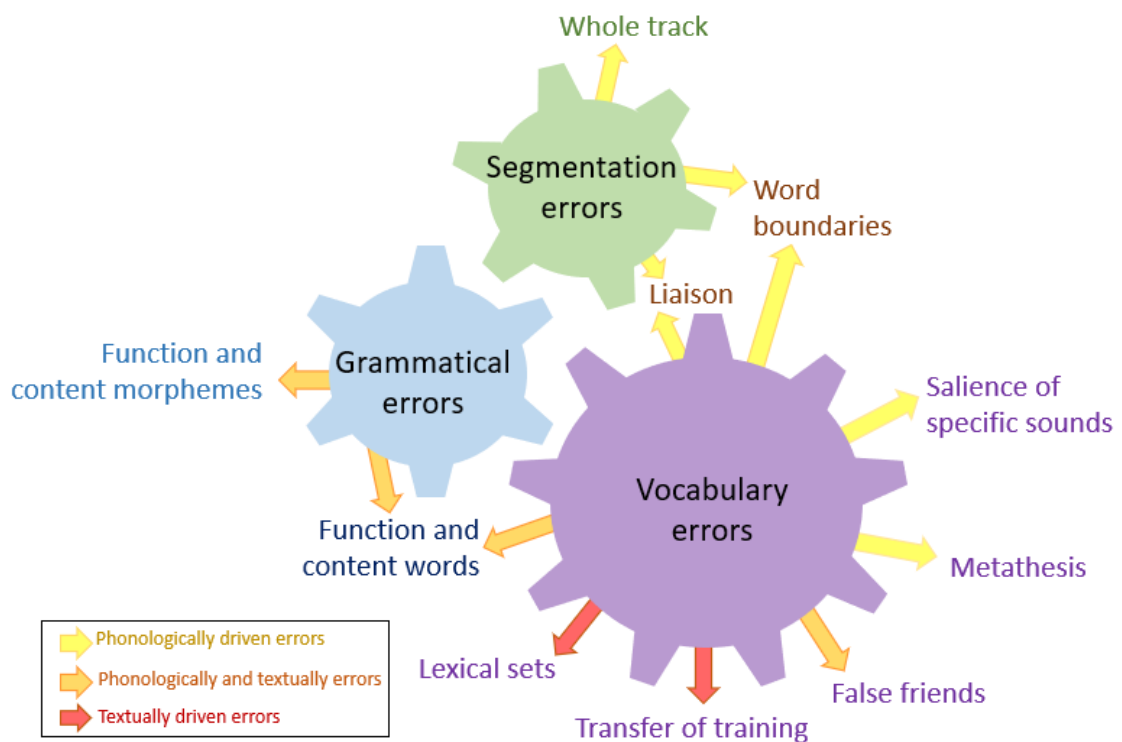


Figure 17: Elements and their categories within this chapter

The intention of both diagram and table is to clearly indicate the commitment to Field’s model (2008b, 2013), while acknowledging both the overlaps and the origin of the errors as either process-driven or textual. Put simply, if a participant had been asked to translate the same text, yet written, could the error have arisen? If the answer is no, it is likely to be phonologically driven (Field’s term is ‘process’), and if the answer is yes, it is likely to be textually driven.

As argued in the literature review, the very experience of listening to an utterance – particularly in a second language – creates a complex dynamic system, whereby the processes are not sequential but iterative and highly interconnected (Cross, 2009; Field, 2008b). From a more philosophical perspective, although there is a need to impose order on the data for analytical purposes, one might also argue that from the learner’s point of view, such attempts at categorisations are immaterial: a misunderstanding or semi-

understanding is just that, with elements of segmental, lexical, and syntactic difficulties and varying levels of compensatory strategy.

This findings chapter does not intend to lead the reader along predetermined reductionist paths based on initial hypothesising; instead, my interpretations and analysis of the data have allowed the emergence of unexpected connections and new perspectives on the experience of L2 listening. To illustrate this at the outset: Time 10, test D3, track 4 consisted of the sentence

sa sœur Martine est très travailleuse

/sa sœʁ martin_ε tʁɛ tʁavajøz/

(‘his sister Martine is very hard-working’),³

which one participant translated as ‘Martin has nice eyes.’ She misunderstood the name, identified the /jøz/ element of *travailleuse* as *yeux* (‘eyes’), and used a filling-in strategy to form a context-appropriate sentence consisting of the two ideas. There are several errors working together which have led to her mistranslation: she has mis-segmented *travailleuse* (segmentation error); she has failed to recognise the stem *travail* (‘work’), of which she has demonstrated knowledge in other contexts (vocabulary error); she has overlooked the *~se* and *~e* feminine agreements (grammatical morpheme errors); and she has overlooked the word *sa sœur* (‘her sister’). What is apparent from this example is the way in which errors of different natures interact with each other to generate often unique

³ A note on style. After both Oxford University Press’s style guide and that of the journal *System*, I have adopted the following style: French words will appear in italics, with their English translations in roman type, in inverted commas, and where appropriate, brackets. Where I am citing a specific word that a participant wrote in the paused translation tests, whether in English or in French, these will also be in inverted commas.

translations, as the listener seizes on whatever they can in order to build some kind of meaning that appears to work in the moment.

While I will be referring to percentages of different types of error (and occasionally the raw numbers), it is crucial to acknowledge that the types of error made by participants depend on the input. As an example, input to which participants were exposed was restricted to the present tense until Time 9, and as such, the ‘wrong tense’ error within the ‘grammar’ category was far less plausible before that time (for full details of the content of each track, plus its English translation, see appendix 15.3). I will be focussing on the data produced by the listening tests which were repeated (see section 3.5.1 for details of which tests when), in some cases up to four times, and discuss how the errors differ over time, and I will then cross-reference with those from any non-repeated tests which demonstrate similar errors. This methodology stands in contrast with a tightly controlled laboratory-based experiment which would control the linguistic input systematically in order to probe specific errors: the longitudinal classroom-based design of the present study did not lend itself to such strictures. The numbers of participants at each time point varies between 68 and 80; for ease of comparison, I have chosen to report percentages of total participants rather than raw numbers. It is also important to note a second figure; that of percentage of attempts. For any section of any test, a certain number of participants wrote nothing at all. A decrease in this number from one test to another indicates progress in that participants who were previously unable to attempt anything have now tried to interpret what they have heard; however this might have the consequence of an increased number of errors, too.

4.1.2. The influence of covid lockdowns

Data collection took place from September 2018 to July 2021, during which time the UK experienced two episodes of lockdown and home-learning, which interfered with the present study. I set out the nature of data collection for each time point in section 3.8.1.

But to summarise briefly, teaching and data collection was as planned up to and including Time 10, and from Times 18 to 20. Between Time 11 and Time 17, participants completed data collection at home and on their own devices.

Contrary to what might be expected, there is no evidence from the paused translation data that the disruption led to either a lack of progress or a reduced engagement. This can be seen by the paused translation scores, over the data collection period are set out in Table 25 and Figure 18, and engagement data (calculated as the percentage of tracks for which participants attempted a response) are in Table 25 and Figure 19.

Table 25: Average scores in paused translation (left), and percentage of attempts (right)

time point	average score	% attempt in test			
time point		test 1	test 2	test3	AVERAGE
T1	36.40	89.7	86.9	79.5	85.4
T2	40.06	83.9	66.2	68.0	72.7
T3	40.26	73.3	67.1	59.1	66.5
T4	44.14	71.3	93.8	61.1	75.4
T5	41.85	82.3	72.5	64.0	73.0
T6	41.59	74.0	67.1	71.3	70.8
T7	47.42	66.1	70.0	71.4	69.2
T8	41.99	80.6	61.3	71.0	71.0
T9	44.62	84.3	65.8	70.4	73.5
T10	47.45	82.2	70.7	66.3	73.1
T11	48.87	78.8	76.3	60.4	71.9
T12	49.07	72.8	77.6	72.7	74.4
T13	49.30	82.5	63.7	74.6	73.6
T14	47.03	77.0	70.2	62.0	69.7
T15	51.66	75.3	69.4	69.2	71.3
T16	51.51	77.4	82.1	72.7	77.4
T17	50.89	78.3	61.9	75.2	71.8
T18	49.36	67.6	89.3	76.7	77.9
T19	52.86	82.3	69.5	72.2	74.7
T20	50.83	69.6	74.0	85.3	76.3

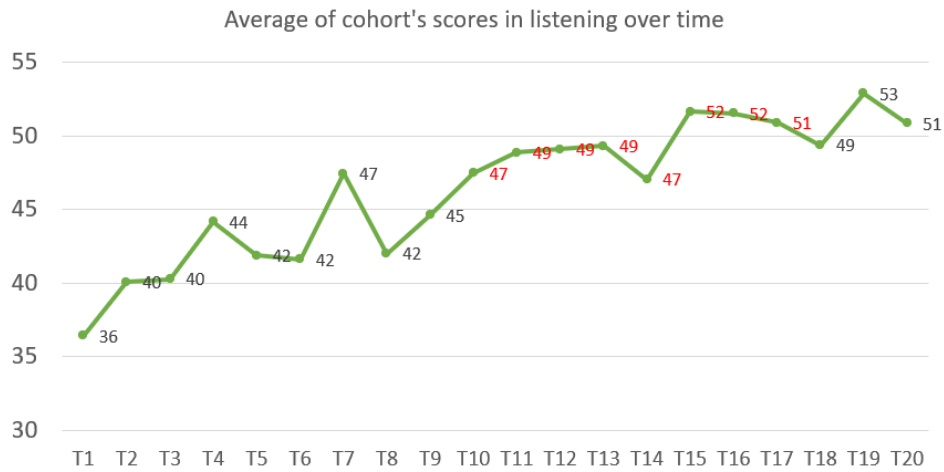


Figure 18: Average cohort listening scores over the data collection period

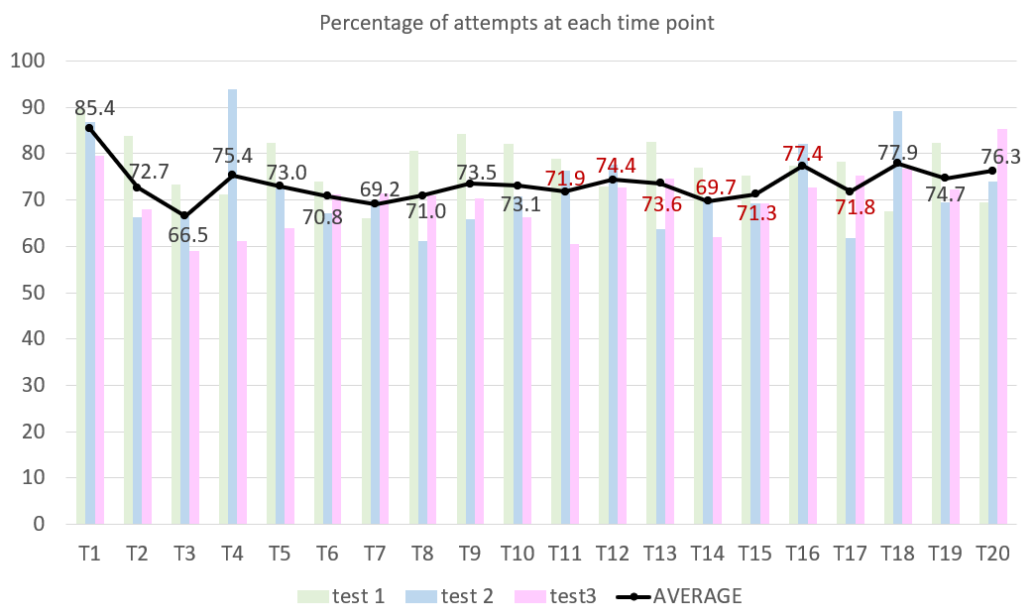


Figure 19: Percentage of attempts at each time point

Nonetheless, for full transparency in this chapter, I mark references to Times 11 to 17 inclusive in red to remind the reader that additional factors might have contributed to the nature of the data submitted. Refer back to Table 22 and Table 23 for full details of location of teaching and testing from March 2020 onwards.

4.1.3. Cautious, ambitious, and reckless test-takers

A further factor complicating data analysis is participants' dispositions. When faced with a track which is only partly understood, some participants guessed at the unclear words,

while others omitted them. Track 5 of test D25 (heard at T9, T13 and T18) illustrates this well: in response to the utterance

Il y a deux bons documentaires sur l'environnement

il i a dɔ bɔ̃ dɔkymɑ̃tɛʁ syʁ lɑ̃virõnmɑ̃

(‘There are two good documentaries about the environment’)

some simply wrote the word ‘documentaries’, while others attempted to form a sentence – mainly ‘I watch documentaries’ or similar: this is illustrated below in

Figure 20.

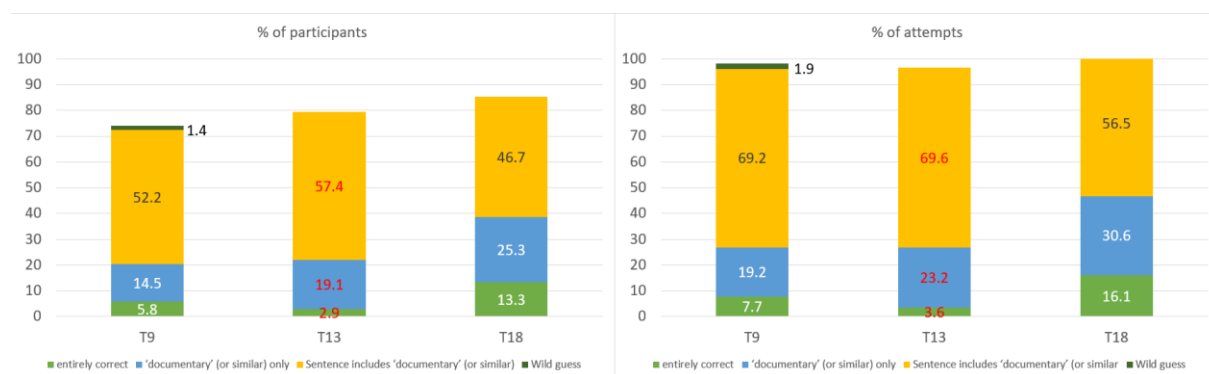


Figure 20: Track 5 of test D25, demonstrating differing dispositions of participants

A further example of caution versus strategic risk-taking – and its change over time – can be noted in track 2 of D47 (heard at T11 and T17). Participants translated *Moi, j’habite dans une maison individuelle* (‘I live in a detached house’) The word *individuelle* caused a problem: many participants wrote ‘I live in a house by myself’, but others avoided the problem word altogether and wrote ‘I live in a house’. The increased inclination to attempt a translation of *individuelle* (whether correct or not) is illustrated in Figure 21.

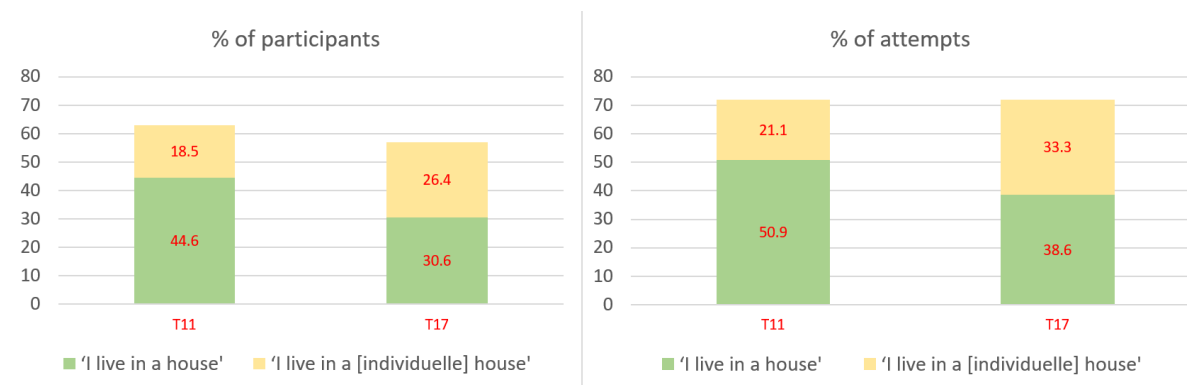


Figure 21: Strategic risk-taking over time

To sum up, participants' approaches to tracks D47 and D25 are discussed here to illustrate a key caveat to this entire chapter: that is, all the findings depend on participants making attempts to translate what they hear. Where participants prefer not to try, what arises is not data but the absence of it, which is not coded, but which is nonetheless of interest in interpreting the findings.

4.1.4. A deficit model of error analysis

As explored in section 3.6, it is important to reiterate that the codebook, which forms the structure of the coming chapter, is built on a deficit model. That is to say, in order to answer the question 'what kinds of errors are made, and how do these errors change over time?', fully correct translations and fully blank entries have not been coded. Rather, all codes relate to an error in the translation.

4.1.5. Codebook summary

A brief summary of the codebook can be found in Table 26. (Full details of the codebook with definitions are in appendix 15.10, and screenshots which illustrate the evolution of the coding are in appendix 15.11).

Table 26: Paused translation errors: NVivo codebook summary

Code	Files	References
Segmentation	59	2722
got end, missed start	52	388
got start, missed end	59	1669
heard start and end; missed middle	32	147
word boundaries not found lead to incorrect meaning	48	502
Vocabulary	60	8691
word missing	60	3412
content word missing	60	1735
function word missing	46	1666
word wrong	60	5279
content word wrong	60	4991
'I like' for a different verb	52	748
linguistically driven false friend	42	495
phonologically driven false friend	58	1057
wrong family-friends word	22	286
wrong pronoun	49	706
function word wrong	42	225
wrong time phrase	30	182
Grammar	58	2346
didn't detect plurals	25	625
hasn't processed verb ending	37	858
incorrect negative added	24	73
incorrect plural given	14	74
missed negative	14	86
wrong tense	32	630

In summary, the codebook contained a total of 13759 coded errors, which are split by percentage as follows: segmentation-related codes accounted for 19.7% of the total, vocabulary 63.2%; grammar 17.1%. Recall that codes were checked with an interrater, who was a fellow languages teacher qualified to Master's level in applied linguistics, who independently coded 10% of the data: initially there was 92% agreement, and disagreements were resolved through discussion.

To recap methodology section 3.6.1, many of the entries provided were double-coded, but each code given was allocated to an entire paused translation cell. Participants' errors interact and even feed off each other: the very first example given in this chapter's introduction illustrates this well. Only coding full sections allowed me to avoid an inbuilt bias towards quantity of vocabulary: in the example cited above there would have been

the potential for four vocabulary references (sister / is / very / hardworking), versus a single code for segmentation, and one for grammar.

I will now explore the errors coded in more detail, following the order in Table 24.

Beginning with the phonological errors which could only have arisen from auditory input, I will discuss errors that have arisen due to problems with word boundaries, liaison, voice-onset, syllable metathesis (misordering), false friends, as well as emergence in the data of the salience of certain sounds. I will examine evidence from the data as a whole, and interrogate the nature of change over time by calling on the repeated tests. I will then discuss a collection of more textual errors. These are vocabulary errors deriving from confusion within lexical sets, transfer of training errors, and the role of content words and function words. The section will be completed with a presentation of grammatically driven morphemic errors.

4.2. Phonologically driven whole-track segmentation difficulties

Here, I will discuss cases in which entire sections of a track were not translated: either the beginning, middle or end of a track was omitted. I will begin by discussing these kinds of cases generally and how this developed over time, and then I will interpret these findings within the context of the existing literature.

Within the dataset 2722 errors were coded to segmentation: detail in Table 27.

Table 27: Errors coded to segmentation

Code	Files	References
Segmentation	59	2722
got end, missed start	52	388
got start, missed end	59	1669
heard start and end; missed middle	32	147
word boundaries not found lead to incorrect meaning	48	502

The first three codes are mutually exclusive: they assumed that the participant correctly translated at least two consecutive words in a track, but then stopped. (If only a single word was missing at the start, end, or middle of the translation of the track, it was coded to ‘word missing’ within the vocabulary category.) It is these three that are discussed in this section.

The ‘word boundaries not found’ code was slightly different, and it refers to cases where a participant has not been able to identify the word boundaries between at least two words – for example within the translation of *ne m’intéresse pas*⁴ /ne mētēʁēs pa/ (‘does not interest me’) a participant gives a translation of ‘spa’; this will be addressed in section 4.3.

4.2.1. Missed sections, and changes over time with missed sections

Participants were much more likely to successfully translate the start of a track, and then stop (1669 references), than the reverse (388 references). One might predict fewer incidents of missing the end as participants’ proficiency increased over the three years. However, when looking at all the repeated tests, a different picture emerges, as can be seen in Figure 22 to Figure 24. The differing numbers of references in each of the tests suggests that this particular code is sensitive to other variables: the lack of discernible pattern points to the interaction between this code and others, and the complex dynamics of the individual moment-to-moment listening experience as well as the complex dynamics of the developmental trajectory of listening proficiency. In fact, to track progress over time, the code is poorly conceived, as it has two different elements to it: the first part of the translation needs to be entirely correct, and then the second part needs to

⁴ track 6 / test D25 / T9, T13, T18

be entirely absent. The low numbers, then, in tests such as E111, D8, D3 and D25 (Figure 22) indicate a certain inherent difficulty with understanding overall, such that only few participants successfully grasped even the start of the tracks.

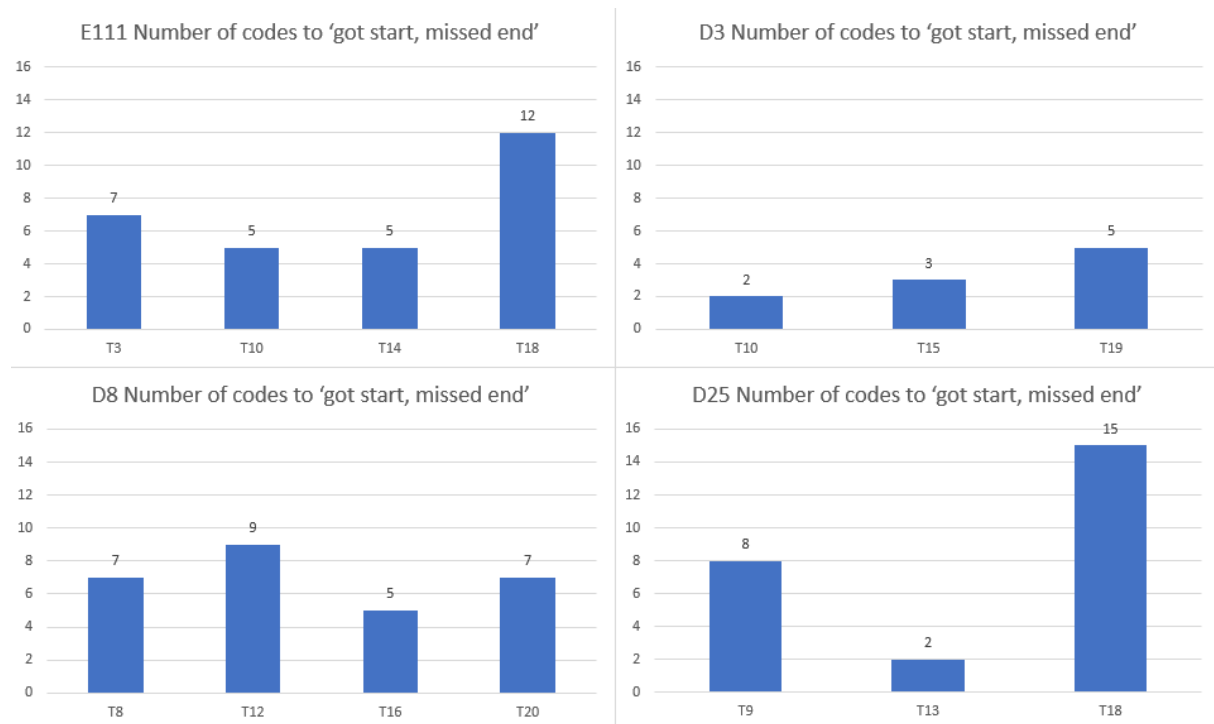


Figure 22: Changes in 'missed the end' with low numbers of references

An argument can also be made that an increase in references to this code, as in test C48 (Figure 23), might suggest that more participants understood the beginning of the track word for word, where before their understanding had been piecemeal: the fact that that this was a difficult track, with attempt percentages between 60 and 70%, also fits with this hypothesis.

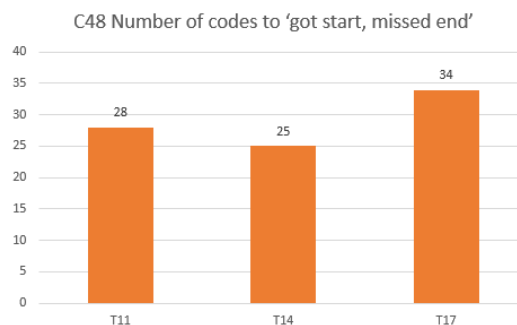


Figure 23: A slight increase in 'got the start, missed the end' code

And by contrast, a decrease in references in the relatively easier tests (Figure 24) could be explained by the fact that more participants understood the entire track, and therefore were not coded to ‘got start, missed end’.

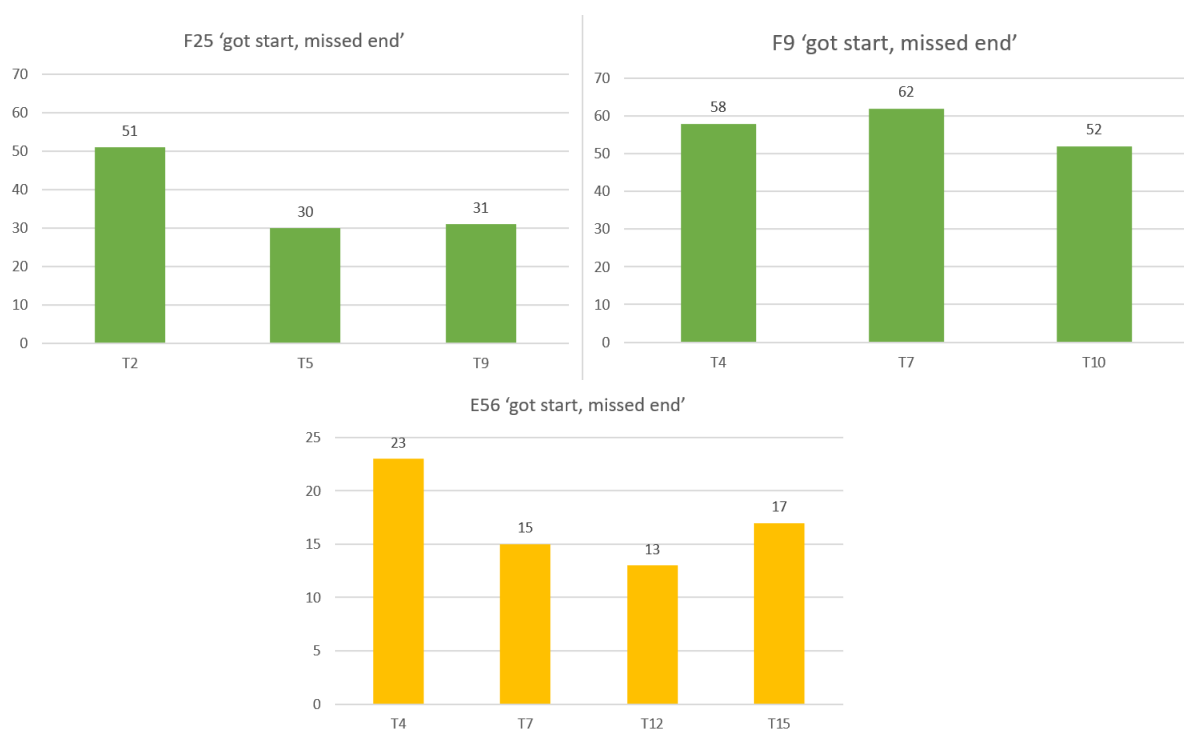


Figure 24: Falling codes for 'got start missed end'

4.3. Phonologically driven vocabulary errors

This section discusses sound-based lexical issues, and is the largest section in this chapter. As a result, there are many more subsections. While again recognising the interplay between types of misunderstanding, it will attempt to draw approximate boundaries around the various categories of error. Key themes will emerge: the persistent interference of English in participants' understanding, often to the detriment of the sense of their translations; that participants' use of context is minimal unless supported by complementary phonological input; and that consonants are far more salient than vowels in the participants' experience. I will discuss errors individually in sections 4.3.1 to 4.3.4, and then explore the developmental trajectory in terms of errors over time in section

4.3.5. All the phonological findings will be put within the context of current scholarship in the Discussion and Analysis chapter.

4.3.1. Word boundary difficulties

I define word boundary difficulties as errors where a mistranslation can be mapped either onto only part of a French word – such as ‘book’ for *beaucoup* /boku/ or straddles two French words, such as ‘thank you’ for *mais c’est*. For my purposes, word boundary difficulties might be driven by the retrieval of either an English or a French word which is homophonous with the French input, however as we will see later, the quantity of the former dominate the findings of the present study.

Where an error was coded to word boundaries, it interacted inevitably with grammatical or vocabulary issues: another example the experience of listening and the interpretation of aural input being a complex dynamic system. In addition to the *travailleuse / yeux* example given in this chapter’s introduction, the following example⁵ illustrates this interplay:

‘Mais c’est beaucoup trop tard’

/mɛ sɛ boku tʁo taʁ/

(‘but it is much too late’)

Although two earlier tracks in the same test had begun with the word *mais* (‘but’), in this track there was a frequent misunderstanding with a total of 20 participants writing ‘thank you’ (*merci* /mɛʁsi/ - phonologically similar to /mɛ sɛ). Some of them had even correctly identified the previous instances of *mais*. Participants’ translation of *mais c’est* as ‘thank you’, is not only a word boundary error, but also a vocabulary error, probably

⁵ track 8 of test D25, T9, T13, T19

exacerbated by the presence of the word *beaucoup*, given that *merci beaucoup* ('thank you very much') is a mainstay of the beginner French curriculum, and is likely to be far more frequently heard than *mais c'est beaucoup*.

The correct detection of word boundaries is a crucial skill in listening to any language at any level, from the beginner learner to the native speaker, and French, with its syllable-timed nature (Tranel, 1987) is particularly challenging in this regard. The data suggest several types of word boundary errors. Firstly, there are word boundary errors whereby an English word is erroneously identified such as the French *bouteille d'eau* /butɛj do/ ('bottle of water') being mistaken for the English word 'potato' /pəteto/. The vast proportion of process errors discussed in this section fall into this category, some neatly, and others initially requiring lateral thinking, (e.g. *plus petit* /ply pəti / ('smallest') as 'puberty') but which shed much light on the struggles of the English beginner learner of French. Secondly there are errors whereby an incorrect French word is identified (e.g. our introductory example of *yeux* for *travailleuse*); errors triggered by context (e.g. *violin* for *vélo* in *j'aime la musique et le vélo* ('I like music and cycling')); and errors where the final syllable and the subsequent initial syllable are the same sound (e.g. *musique classique* /mysik klasik/ translated as 'music lesson' instead of 'classical music'. This example could have a range of roots which also point to Anglo-French cognates: class or the French cognate *classe* /klas/, or lesson or the French cognate *leçon* /ləsɔ̃/.

The illusion of English words: French-English boundary problems create pseudocognates

By far the commonest category of word boundary issues are where a participant formed their translation around an English word which sounds similar to the French input. This type of error often resulted in a nonsensical translation and speaks to the finding which arises in many different contexts within this chapter, which is that of 'sound trumps sense'.

To illustrate this point, there were frequent incidents within the data whereby participants latched onto a particular sound within the French input and gave it English meaning despite contextual irrelevance. (Recall that participants were told the context of each passage before beginning to listen – see

Figure 8.) For example, in test C48⁶, participants were asked to listen to someone describing their home, and track 2 consisted of the phrase:

il y a plein d'espace pour tous mes vêtements

/il j_a plɛ̃ dɛspas puʁ tu me vɛtmɑ̃/

(‘there is plenty of room for all my clothes’)

Over the four time points, on 18 occasions participants missegmented *espace pour* /**espas puʁ**/ and gave ‘passport’ as their translation. Furthermore, in the same test,

mais la salle de bains est loin de la chambre

/mɛ la sal **də bɛ̃** ɛ lwɛ̃ də la ʃɑ̃brɔ/

(‘but the bathroom is far from the bedroom’)

ended up with 13 references to Dubai, and the track

Notre chambre est au quatrième étage

/nɔʁkɑ ʃɑ̃brɔ ɛt **o** katʁiɛm ɛtaʒ/

(‘our room is on the fourth floor’)

⁶ T11, T14, T17 and T20

generated two translations of ‘cats’ and one of ‘pets’ and four more of ‘four’ (French *quatre* /katʁ/ rather than fourth).

Other examples include ‘spa’ for *ne m’intéresse pas* as in section 4.2, which arose five times at Time 9, four at **Time 13** and eight at Time 18, and a large collection of one-off mis-segmentations, such as ‘map sir’ for *ma petite sœur*⁷ /**ma** pətɪt sœʁ/; ‘purdah’ for *pas de*⁸ /**pa** də/; both ‘my sir prefer sir’ and ‘I prefer my professor’ for *ma sœur préférée*⁹ /ma sœʁ pʁɛfɛʁɛ/; ‘books’ for *beaucoup*¹⁰ /**boku**/; ‘doesn’t have a lamp’ for *je ne l’aime pas*¹¹ /ʒə nə **lɛm pa**/; ‘glue’ for *avec lui* /avɛk **lyi**/; and ‘igloo’ for *rigolo*¹² /**ɪgolo**/; as well as, for *ski nautique*¹³ /**ski** notik/ ‘skinning people’.

Naughty singers and things on the ground floor: French-French boundary problems

Within the dataset of 442 word boundary errors, there are only 40 which evidence knowledge of an alternative French word with no English near-cognate, one of which is the example given in the introduction to this chapter, whereby the syllable /jøz/ in the word *travailleuse* /tʁavajøz/ is taken to mean ‘eyes’ (yeux /jø/). The paucity of errors of this nature speaks again to the theme that ‘sound trumps sense’ in that in the main participants appear to be driven by attempts to assign English meaning to French sounds, as discussed in the section above.

But where French to French boundary problems have arisen, it demonstrates the overlap of many errors, as discussed in the introductory example, whereby segmentation, lexical and syntactic errors interact to create unique misunderstandings or attempts to make

⁷ track 4 test C2 **T15**

⁸ track 9 test D47 **T17**

⁹ track 8, test E3, T8

¹⁰ track 4, test St1, T2

¹¹ track 5, test E56, **T13**

¹² tracks 2 and 4, test C55, **T13**

¹³ track 2, test St342A, T20

sense of the input. Another example can be seen in the liaison section 4.3.1 where /ty/, meaning ‘you’ is derived from /paʁsə kə sɛt_ ytil/. All examples are given in Table 28. Where tests are repeated, I have indicated with a * if it is the same participant at each time point.

Table 28: All examples of French-derived missegmentations

Time	Track and test	French input	IPA	English mistranslation	Explanation	# incidences
5	Tr2 E3	<i>méchante</i>	/mɛʃɑ̃t/	sings	= <i>chante</i> //ʃɑ̃t/	1
9	Tr8 D25	<i>mais c'est beaucoup</i>	/mɛ sɛ boku/	thanks a lot	= <i>merci beaucoup</i> /mɛʁsi boku/	6 (2*)
10	Tr4 E111	<i>c'est utile</i>	/sɛt_ ytil/	it is you	= <i>c'est tu</i> /sɛ_ty/ (grammatically incorrect)	3
10	Tr4 D3	<i>travailleuse</i>	/tʁavajøz/	eyes	= <i>yeux</i> /jøz/	1
11	Tr3 D47	<i>rez-de-chaussée</i>	/ʁɛdʒoz/	things	= <i>choses</i> //ʒoz/	1*
12	Tr2 C17	<i>ma tartine</i>	/mataʁtin/	morning	= <i>matin</i> /matɛ̃/	1
13	Tr1 C55	<i>je m'entends</i>	/ʒəmɑ̃tɑ̃/	I like	= <i>j'aime</i> /ʒɛm/	2
13	Tr8 D25	<i>mais c'est beaucoup</i>	/mɛ sɛ boku/	thanks a lot	= <i>merci beaucoup</i> /mɛʁsi boku/	4 (2*)
14	Tr4 E111	<i>c'est utile</i>	/sɛt_ ytil/	it is you	= <i>c'est tu</i> /sɛ_ty/ (grammatically incorrect)	1
15	Tr4 D3	<i>travailleuse</i>	/tʁavajøz/	eyes	= <i>yeux</i> /jøz/	1
16	Tr2 E3	<i>méchante</i>	/mɛʃɑ̃t/	sings	= <i>chante</i> //ʃɑ̃t/	1
18	Tr4 E111	<i>c'est utile</i>	/sɛt_ ytil/	it is you	= <i>c'est tu</i> /sɛ_ty/ (grammatically incorrect)	1
18	Tr8 D25	<i>mais c'est beaucoup</i>	/mɛ sɛ boku/	thanks a lot	= <i>merci beaucoup</i> /mɛʁsi boku/	10 (2*)
18	Tr3 St3AP	<i>mon frère aime</i>	/mɔ̃n fʁɛʁ ɛm/	my brother and	= <i>mon frère et</i> /mɔ̃n fʁɛʁ ɛ/	5
19	Tr4 D3	<i>travailleuse</i>	/tʁavajøz/	eyes	= <i>yeux</i> /jøz/	1
19	Tr3 D47	<i>rez-de-chaussée</i>	/ʁɛdʒoz/	things	= <i>choses</i> //ʒoz/	1*

Note also that the 40 errors are disproportionately skewed towards later timepoints, demonstrating the influence of an increased vocabulary as participants develop. I considered whether carrying out tests at home between **Time 11** and **Time 17** might have been a factor, but given the continuation of the problems from Time 18 onwards, conclude that this is unlikely.

I say 'bouteille d'eau', you write 'potato': contextually feasible translations

Within the category of word boundary issues there are also errors which are much more contextually feasible than the ‘sound trumps sense’ examples given in the previous

paragraph. There are three pertinent examples of this within the dataset. The first of these¹⁴ was found in a passage about what to order in a café, where participants heard:

mais je voudrais bien une bouteille d'eau

/mɛ ʒə vudʁɛ bjɛ̃ yn butɛj do/

'but I'd quite like a bottle of water'.

In this test there were 78 participants of which 56.4% (44 participants) attempted a response, and 15 gave the translation 'a potato' (19.2% of all the participants, and 34.1% of those who attempted a response), compared with only two participants who mentioned bottles of water and another translation of one bottle with no mention of water. Similarly, in the same test, track 10 referred to

une saucisse avec des frites

/yn sosis avɛk de fʁit/

'a sausage with chips'

and seven out of 57 participants (9% of total; 12.3% of those who attempted a translation) translated *saucisse* as 'sauce'.

A further example of contextual feasibility can be found in a passage¹⁵ about activities which teenagers are allowed to do, containing the line,

mais je ne peux pas rentrer seule

/mɛ ʒə nə pø pa ʁɑ̃tʁɛ sœl/

'but I am not allowed to come home alone'.

¹⁴ track 12, test C41, T16

¹⁵ track 7 of test St3AP, T19

64 participants participated in this test, of which 35 (54.7%) attempted a translation of this track. Eleven (17.2% of total; 31.4% of those who attempted) missegmented /pa kãtœ/ to end up with ‘parents’ /pa kã/, compared with six (9.4% / 17.1%) whose answer referred to going out alone. None grasped *rentrer* as ‘to come home’.

Within the dataset there are many other examples of one-off contextually appropriate errors of segmentation, such as ‘I watch colour TV’ for *je regarde beaucoup la télé*¹⁶, /ʒœ kœgãd boku la tele/ (‘I watch the television a lot) and ‘I go to tech’ for *je vais à la bibliothèque*¹⁷ /ʒœ ve a la bibliotœk/ (‘I go to the library’) (where ‘tech’ is a school subject and the library referenced is also in relation to school).

Le musée est formidable: some particularly salient sounds

Words containing two key French sounds - /mys/ and /abl/ - were frequently missegmented and mistranslated (hence also coded to vocabulary). Participants were exposed to the sound /abl/ on 20 occasions¹⁸ during the data collection period. The word *portable* (mobile phone) caused few problems, as can be seen in Figure 25.

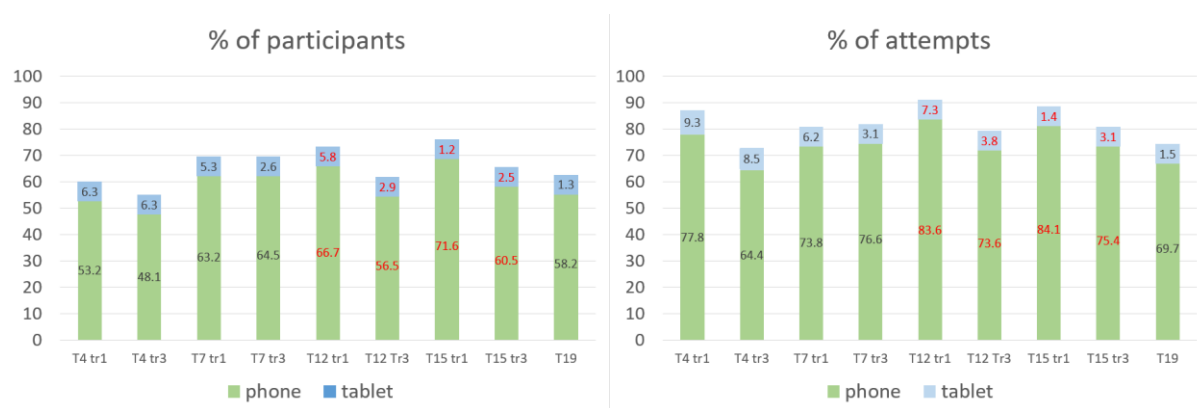


Figure 25: Translations of the sound /abl/ when given the word portable

¹⁶ track 1, test St22A, T8

¹⁷ track 7, test E111, T3, T10, T14 T18

¹⁸ ‘Portable’ x 9 at tr1 and tr3 of E26 at T4 T7 T12 T15, plus tr1 of St3151A T19; ‘formidable’ x7 (tr2 of D3 at T10, T15, T19, and tr1 of E3 at T5 T8 T11 and T16; ‘adorable’ x4 (tr4 test E3 T5 T8 T11 T16)

The majority of participants who attempted a translation did so correctly, although we can see a small number who missegmented – yet within a lexical set of ‘digital devices’ to come up with a translation of ‘tablet’.

Yet ideas of phones and tablets also predominated in other /abl/ words, such as *formidable*, (Figure 26) where a small number of participants wrote ‘phone’, ‘tablet’ or ‘table’. It is notable here that the translations have moved beyond the ‘digital devices’ lexical set which was cited above, suggesting the salience of the /abl/ sound was the deciding factor in participants’ choice of translation rather than a technology schema, and that the similarity of *portable* and ‘tablet’ is a red herring. In the ‘consonant pairs’ section of part 4.3.3, I will return to these cases and discuss whether a confusion between /tbl/ and /dabl/, might also have led to or exacerbated these difficulties.



Figure 26: Attempts at translating formidable in tests E3 and D3

By contrast, translations of *adorable*, (Figure 27) presented only minimal problems – ‘tablets’ hardly featured – and there was no evidence of other segmentation difficulty. The presence of translations such as ‘horrible’ and ‘terrible’ suggest participants perceived not /abl/ but /ʁ~bl/. I will return to this hypothesis in the section 4.3.3 (consonants and vowels) below, and discuss the pedagogical implications in a future chapter.

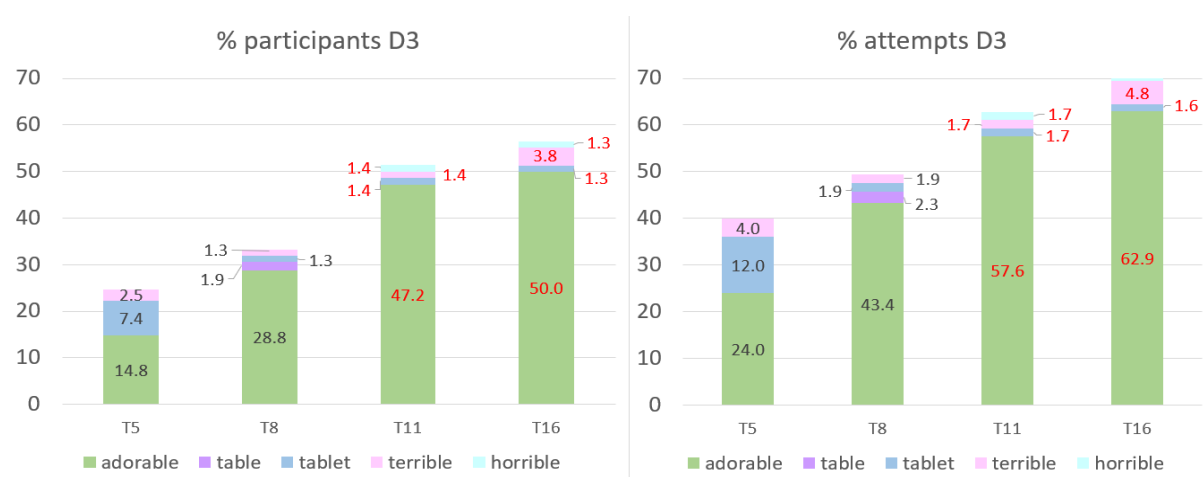


Figure 27: Translations of *adorable* in test E3

Translations of all words containing the sound /mys/ tended to funnel towards a translation of ‘music’. Where *musique* occurs within the dataset¹⁹ (16 times), there is no evidence of missegmentation or mistranslation. The word *amusant* causes some level of difficulty with a total of six incidences of missegmentation and/or mistranslation of ‘music’ and three of ‘museum’. But *amuser* or *amusé* /amyse/ presents far more difficulties – this is no surprise when one considers that it contains within it a complete homophone with the French word for ‘museum’ /myse/. Errors are given below in Figure

¹⁹ ‘Musique’ appears as follows: tr5 of test AS7 at T1 and T4; tr7 of test F25 at T2, T5, T9; tr4 of test E26 at T4 T7 T12 T15; Tr2 of test St212A at T8; tr8 of test D8 at T8 T12 T16 and T20, and tr9 of test C48 at T11 and T17.

‘Amusant’ appears as follows: tr4 of test St2 at T8; tr3 of test D3 at T10, T15, T19; Tr3 of test St42a at T20.

‘Amuser’ or ‘amusé’ appears as follows: tr5 of test D3 at T10, T15, T19; tr 11 of test St3AP at T18.

28. Note that Time 10, Time 15 and Time 19 data relate to test D3; Time 18 data relate to test St3AP, and the relative success in recognising the word in test St3AP speaks to the interdependence of factors when participants are struggling to make sense of aural input.

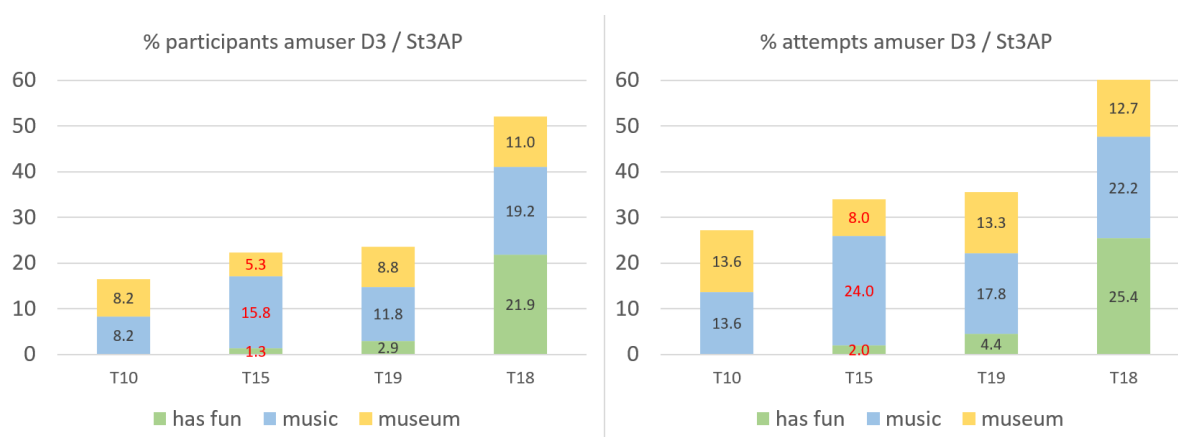


Figure 28: Missegmentation of amuser

Intuitive tutus are not useful: Liaison

In French, liaison is the practice of pronouncing a letter which would ordinarily be silent in order to avoid two vowel sounds occurring next to each other in connected speech.

This can confuse beginner learners who then fail to hear a word they are expecting, as it appears to begin with a different sound. As with many of the cases explored in this section, there can be more than one explanation for the error: in the following example, the mistranslation is also contextually feasible (see ‘I say *bouteille d’eau*’ in section 4.3.1, above).

In a passage with a ‘holidays’ theme²⁰, participants heard:

²⁰ track 7 of test St3AP at T18

On est allés au bord de la mer

/ɔ̃n_ɛt_ale o bɔʁ də la mɛʁ/

(‘we went to the seaside’)

But out of 73 participants, 51 of which attempted a response, only three (4.1% of total; 5.9% of attempts) gave ‘we went’ for *on est allés*, compared with 30 (41.1%; 58.8%) who wrote ‘in Italy’. It is also notable that two of the three participants who translated *on est allés* correctly also included reference to Italy; in other words, they translated /ɔ̃n_ɛt_ale/ twice.

Key errors arising from misunderstood liaison were also apparent in the track²¹

parce que c’est utile

/paʁsə kə set_ytil/

(‘because it is useful’)

where participants veered towards a missegmentation of /set_ytil/ to give /ty/. This led to an attempt at forming a meaning surrounding a word with the sound /ty/ or /tyt/ – mentions of the number ‘two’, a neologism ‘**tutile**’, ‘**tutu**’, ‘**intuitive**’, ‘**factual**’, ‘**stupid**’ and ‘you’ (*tu* in French). Where participants opted for *tu* (you) as the key word around which to form meaning, syntactic logic forced them to follow this with the word ‘are’, and it then followed that the word *c’est* (‘it is’) could not be included in the translation (e.g. ‘because you are [adj]'). By contrast, when the participants’ key word was ‘factual’, ‘stupid’, ‘intuitive’ or even the made-up ‘tutile’, ‘it is’ could be included in the translation quite naturally (e.g. ‘because it is factual’). A third option for participants’ key words were ‘two’ and ‘tutu’, which forced the formation of the verb ‘to have’ (e.g. because I

²¹ track 4, test E111, T3, T10, T14, T18

have a tutu). This is set out in Figure 29. A key point in the ‘sound overrides sense’ argument emerges from these data. Participants’ translations may not make logical sense, but their translations are consistently grammatical in English.

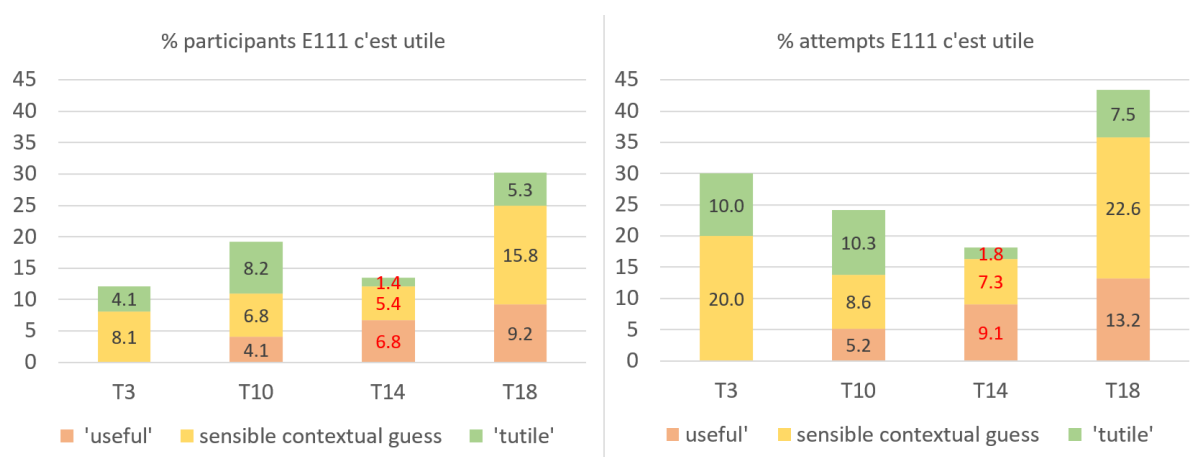


Figure 29: Dealing with liaison in c'est utile

4.3.2. Word boundary difficulties over time

Like the hypothesis posed above in section 4.2.1 regarding changes over time with missing sections, one might initially expect to simply see fewer word boundary difficulties as the data collection progressed, but the reality is more complex. Therefore, we now need to ask whether the nature of the difficulties has changed and what factors might be implicated in such a change.

To answer these questions, I formed three groups from the repeated listening tests: tests where there is an increase in word boundary error references, ones where there is a decrease, and ones where there is no change. Each of these is discussed in turn below.

Tests where word boundary errors increased

The only tests in which word boundary errors appeared to increase over time were tests D25 and C48. For C48, it appears that the reason for the increase in recorded errors is an increase in recorded attempts made – see Table 29.

Table 29: Word boundary errors and attempts at translation for test C48

C48	T11	T14	T17	T20
# word boundary errors referenced	8	5	7	19
% attempts made	60.4	62	61.9	69.6

C48 was one of the hardest tests within the project, (see appendix 15.14 for full data), and even the erroneous detection of the cognate *passeport* for *espace pour* was beyond most participants. Such an inability implies that there is a stage in auditory language processing that precedes the learner finding even familiar L1 sounds or words within an L2 input: that of sheer overwhelm. The period between **Time 17** and Time 20 was a little more than four months, but as far as test C48 is concerned, the increase in errors is indicative of an increase in willingness to attempt the task (61.9% at **Time 17**, and 69.6% at Time 20), possibly linked returning to in-school test-taking, or to motivation and self-efficacy. This will be investigated further in the next findings chapter which probes the relationships between the different subsystems contributing to L2 listening.

The key segmentation issues in D25 are given in Table 30, where we can see that there is an increase in absolute number of errors but also in number of attempts: in other words, just as with test C48, by Time 18 more participants are willing to have a go, and risk being wrong.

Table 30: Principal segmentation errors recorded from D25

D25	T9	T13	T18
# participants	69	68	75
Total recorded segmentation errors	15	15	25
# attempts at track 4	59	59	67
Track 4 'Derek' for 'Frédéric'	3	4	2
# attempts at track 6	27	9	27
Track 6 'spa' for 'ne m'intéresse pas'	5	4	9
# attempts at track 7	57	48	57
'cinema' for 'dessins animés'	1	2	4
# attempts at track 8	37	31	37
'Merci beaucoup' for 'mais c'est beaucoup'	6	4	10

Tests where word boundary errors decreased

In contrast with the increase in referenced errors for tests C48 and D25, Table 31 shows the three repeated tests in which the number of word boundary errors fell. These are tests F9, E3 and E26. I suggested above that one reason for an increase in boundary errors could have been an increase in effort, but we can see in Table 31 that, at the very least, the reverse pattern is not visible here: effort levels appear to stay the same or increase slightly along with the decrease in boundary errors.

The theme of F9 was sport-related hobbies, and the key item which presented a segmentation challenge was the phrase *pour faire* /puʁ fɛʁ/ ('in order to do'), which was sometimes misunderstood as *préfère*²²/pʁɛfɛʁ/ (9, 6 and 4 times respectively), plus a single mistranslation of 'professor' at Time 7. All other examples of missegmentation here are unique: the word *l'équitation* /lekitasjɔ̃/ ('horseriding') was missegmented into 'lake' once at Time 4 and Twice at time 10, and *au bord de la mer* /o bɔʁ də la mɛʁ/ ('at the seaside') became 'aubergine' once at Time 4. The familiar topic might have been instrumental in participants' willingness to attempt the task.

Table 31: Tests which indicate a decrease in word boundary errors

E26	T4	T7	T12	T15
# word boundary errors referenced	11	6	8	3
% attempts made	61.1	71.4	72.8	75.3

E3	T5	T8	T11	T16
# word boundary errors referenced	20	10	8	11
% attempts made	64.0	61.3	78.8	77.4

F9	T4	T7	T10
# word boundary errors referenced	11	7	6
% attempts made	71.3	70	70.7

²² 9x at T4; 6x at T7 and 4x at T10

E26 and E3 have already been discussed above, with data given in Figure 25 to Figure 27, and the findings speak to the same idea: frequent vocabulary such as *portable* (‘mobile phone’) becomes increasingly familiar and therefore, over time, less likely to pose a segmentation problem.

Tests where word boundary errors remained unchanged

The final category of tests, set out in Table 32, comprises those with no change in word boundary errors. D3, whose key segmentation challenges were the words *formidable* – an infrequent word – and *amuser* – a difficult near-homophone discussed above in section ‘musique amusant amuser musée’ in 4.3.1– see Figure 28.

Table 32: Tests which indicate no or minimal change in word boundary errors

D3	T10	T15	T19	
# word boundary errors referenced	21	22	19	
% attempts made	66.3	69.4	69.5	
E111	T3	T10	T14	T18
# word boundary errors referenced	21	22	15	21
% attempts made	59.1	82.2	77	77.3

Errors arising in test E111 have been discussed above within the context of liaison in section 4.3.1 (*parce que c’est utile*). Another low frequency word – *utile* (‘useful’) – might be a contributory factor, recognition of which does not improve over the data collection period. To conclude, then; it is not word boundary issues *per se* which improve over time, but rather change (or lack of it) points to the interaction of segmentation skills and vocabulary knowledge. More frequently encountered words become easier to recognise within the stream of speech, but rarer words continue to cause problems. In other words, this suggests that vocabulary knowledge drives segmentation skills rather than the reverse.

4.3.3. Further lexico-phonological errors

Within the errors coded to vocabulary, there is a range of other mistakes which are phonological in origin, although they relate more closely (although not always entirely) to complete words than to segmentation problems. The coming sections refer to the linguistically driven vocabulary issues of ‘salience of specific sounds’ and ‘metathesis’ as identified in Table 24.

It appears from the data that participants in the present study prefer to form meaning via consonants rather than vowels, but continues to speak to the hypothesis which is emerging from the data that the *sound* of what is heard drives the participants’ interpretation as much as contextual feasibility; and any chance that both sound and context could work in tandem is seized upon.

‘Puberty in my family is very funny’: Consonant pairs

In terms of voice onset timing in pairs of voiced/voiceless consonants, French and English listeners would disagree about the point at which the /k/ sound becomes /g/ (Cho, Whalen, & Docherty, 2019). The same can be said of the pair /p/ /b/ and /t/ /d/. As a rule, it is reasonable to state that what is perceived as voiced by a French listener would in some cases sound unvoiced to English ears. In technical terms, this is due to a difference in timing of the onset of voicing between French and English (ibid).

These three sets of consonants caused confusion throughout the dataset. The /g/ /k/ pair suggested a possibility of interference of voice onset timing, whereby *pain grillé*²³ /pɛ̃ ɡʁijɛ/ (‘toast’) was translated as ‘pancake’ by 22 of the 60 participants (36.7%) who

²³ track 5 test C17 T12

attempted a translation – an example of context and sound working together to bring listeners to an incorrect yet feasible answer. Yet participants did not always adhere to the feasible: in a passage about eating out²⁴, the word *gateaux* /gato/ (‘cakes’) was translated in four cases (78 participants, 48 attempts) as ‘cat’ (plus a dog) alongside the 31 correct translations of ‘cake’²⁵.

There is further evidence for the interference of voice onset timing between /p/ and /b/ with the example of *plus petit*²⁶/ply pəti/ (‘smallest’), translated by a different single participant as ‘puberty’ at each time point; between /t/ and /d/ in *formidable* becoming ‘table’ (see Figure 26), and both pairs *bouteille d’eau* /butɛj do/ as ‘potato’ by 15 out of 44 participants, discussed above in section 4.3.1.

Yet there are also cases where participants have heard an unvoiced French consonant and given it a voiced translation: this can be seen in where *parce que* /pɑksə kə/ (‘because’) is translated as ‘basket’ on nine occasions²⁷ and ‘j’attends’ /zatɑ̃/ (‘I wait for’) is translated as ‘I love’ (French ‘j’adore /ʒadɔʁ /) 20 times²⁸.

Within the full dataset there are 80 examples whereby a French voiced consonant is mistranslated as an English unvoiced consonant, which is consistent with the phonology of the two languages. Yet there are also 29 examples of the opposite, suggesting that perhaps consonant pairs more generally might be what is confusing the beginner listener rather than the minutiae of levels of voicing.

²⁴ track 5 test C17 T12

²⁵ 5.1% / 8.3% of attempts for cat; cake = 39.7% / 64.6%

²⁶ track 3 test D3 T10, T15, T19

²⁷ test E111 (T3 T10 T14 T18)

²⁸ (T3x2 T10x10 T14x6 T18x2)

Du camping à la campagne avec des copains, et du vélo en ville : consonants and vowels

There is also evidence to suggest that vowels are largely disregarded: the upshot of this is that pairs of words whose consonants match but whose vowels differ were problematic.

This was particularly the case when the are very likely to appear in the same passage, or a learner’s schema, such as the words *camping* / kãpiŋ / (‘camping’ or ‘campsite’), *campagne* / kãpaŋ/ (‘countryside’) and *copains* / kɔpɛ̃ / (‘friends’).

However, as we have already seen in other contexts, one might also argue that the root of such mistranslations is an interaction with word frequency and vocabulary knowledge: in the six tracks in which the word *copain* (‘friend’) appears within the dataset, it is never mistranslated. However, participants tended to assign a meaning of ‘copain’ to other words with the /k-p/ sound. Figure 30 illustrates, for example, participants’ mistranslations of the word *campagne*²⁹.

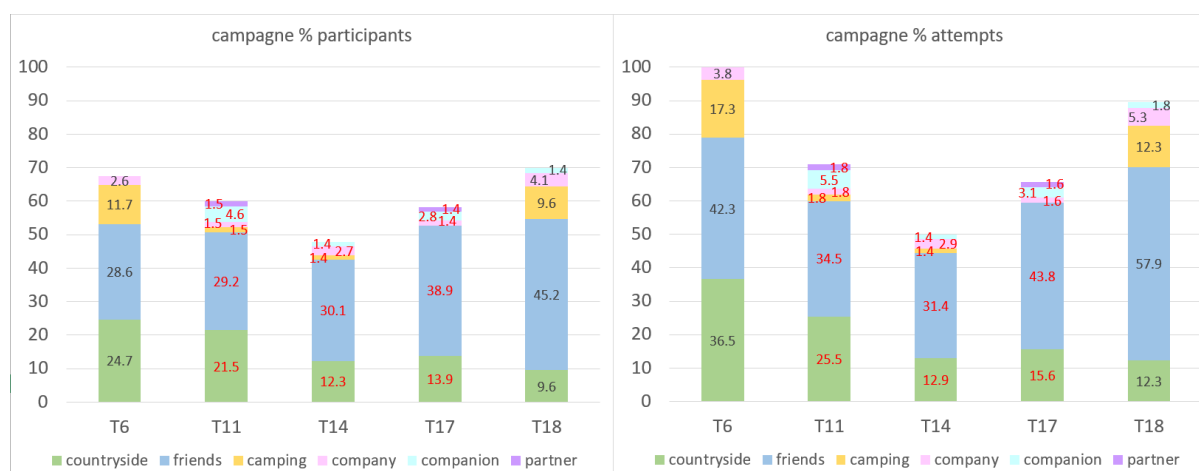


Figure 30: Mistranslations of *campagne* over time

However, the only inclusion in the dataset of the word *camping* was within the phrase *on fera du camping* (‘we will go camping’)³⁰, and 53.6% (30/56) participants who attempted

²⁹ T6 in test G32, T11 in test D47, T14 in test D31, T17 in test D47, T18 in test St3AP

³⁰ T18 (track5, test St3AP)

a translation correctly translated the word. But even at Time 18 the draw of the *copain* is strong, and a further 26.8% (15/56) translated it as ‘friends’. As if to illustrate yet again the complexity of interaction between different elements of aural input, however, it could also be possible that the word *fera* /fəʁa/ (‘will do’) was the origin of the choice of the word ‘friend’: none mentioned both camping *and* friends, which perhaps would have happened had *fera* been understood as ‘friend’, but as we have seen time and again with the analysis of the data, influences leading participants to give specific translations are manifold. Findings chapter 3 will allow me to examine such influence in more detail with the analysis of the case studies’ stimulated recall data.

Participants’ focus on consonants at the expense of vowels is also seen with the efforts to translate the word *vélo* (‘bike’ / ‘cycling’), although unlike the *copain* example, here listeners lack the opportunity to map what they hear onto a high-frequency word. This lack appears to have led to more accurate translations. However, again there is an interaction between phonological knowledge and vocabulary knowledge, as illustrated in Figure 31: it is quite possible that the drop in correct identification of the word *vélo* at Time 12 is because of an increase in general vocabulary size – at the outset, there are few competitors for the word. But as time progresses, participants have a growing vocabulary and therefore a growing set of possible matches for a /v-l/ words.

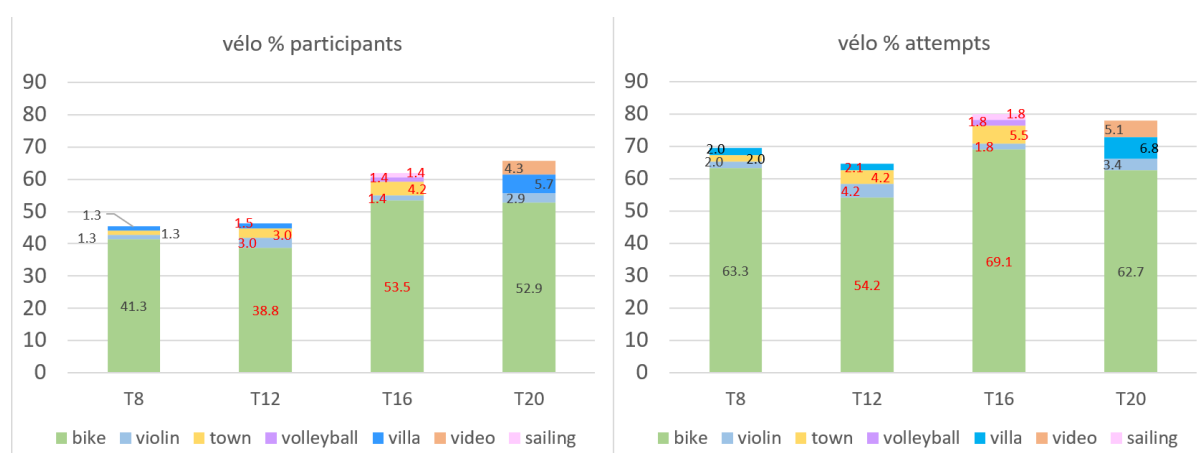


Figure 31: Translations of ‘vélo’ test D8

The interaction between phonological knowledge and context was also apparent in a passage about hobbies³¹: *vélo* appeared in the track 5 *j'aime le vélo et le foot* ('I like cycling and football') and later in track 8 *j'aime le vélo et la musique* ('I like cycling and music'). We see a small increase in favour of the translation 'violin' in track 8, as can be seen in Figure 32, below, as well as a reduction in interference at Time 4 compared with Time 1.

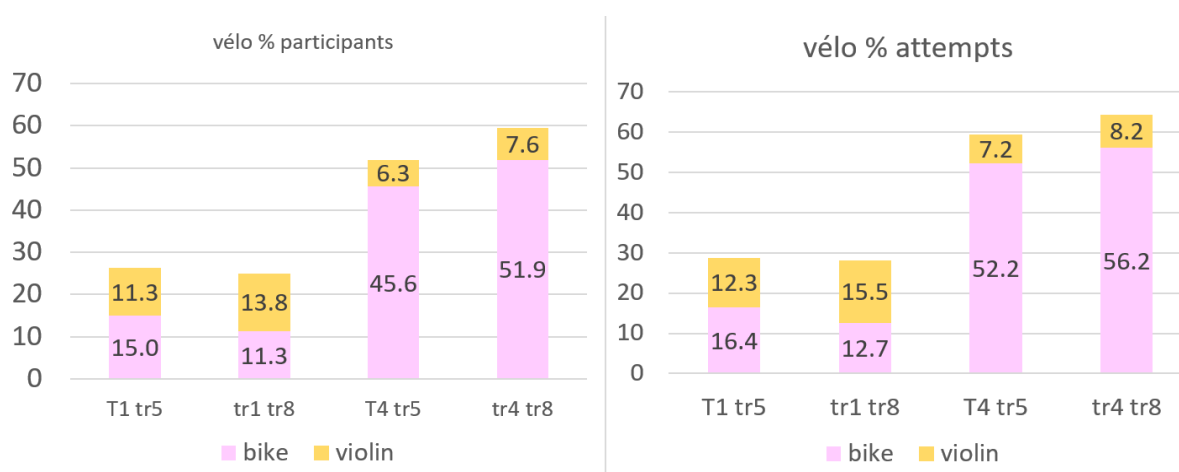


Figure 32: Contextual interference for translation of *vélo* at T1 and T4

Consonant salience confusion is not wholly confined to consonant-vowel-consonant words though: in the phrase *je ne mange rien avant l'école*³² (*I eat nothing before school*), the /ko/ sound appears salient in translations, which include 'I eat **curd**', 'I eat **cod**', 'I eat **bacon**' and 'I eat **cooked things**'. Similarly in *J'ai faim*³³, the /fa/ sound drives mistranslations such 'I do' (*je fais*), 'my family', and 'fun' and one translation of *mon frère Dominique ne cesse jamais de parler*³⁴ ('my brother Dominic never stops talking') was 'my brother Dominic never speaks unnecessarily', homing in on the /ses/ sound.

³¹ Test AS7 which took place at T1 and T4

³² track 8 test C17 Time 12

³³ track 1 test C41 T16

³⁴ track 1 test 55 T13

Un jardin d'argent: Metathesis

There are 21 examples within the dataset of syllable metathesis (the process whereby syllables are misordered) and they are given in Table 33.

Table 33: Instances of syllable metathesis across the dataset

Input	IPA	Metathesised French word	Metathesised mistranslation
<i>dessins animés</i>	/ desɛ̃z anime/	<i>cinéma</i>	cinema
<i>natation</i>	/ natasjɔ̃/	<i>tennis</i>	tennis
<i>d'argent</i>	/ daʁzɑ̃/	<i>jardin</i>	garden
<i>calme et le repos</i>	/kalmeləʁpɔ/	<i>caramel</i>	caramel
<i>recréation</i>	/ ʁəkʁeasjɔ̃/	<i>correction</i>	correction
<i>sœur préférée</i>	/sœʁ pʁefɛʁe/	<i>professeur</i>	teacher
<i>gentil</i>	/zɑ̃ti/	<i>intelligent</i>	intelligent
<i>sœur</i>	/sœʁ /	<i>russe</i>	Russian

These cases appear across the dataset from Time 2 to Time 18, and participants of differing levels of proficiency and motivation. Although this is a tiny proportion of the number of words which have been mistranslated, the naturalistic nature of the input was not designed deliberately to induce such errors. Yet the very existence of errors of this nature continues to speak to the notion that the *sound* of the input overrides meaning, and a salience of consonants, albeit now not necessarily in the right place. Some examples cross word boundaries, such as the /z anime/ section of the word *dessins animés* ('cartoons') being translated as 'cinema', which accounts for seven of the cases within the dataset.

Doubling up

Another rare but noteworthy phenomenon within the dataset is that of twice translating the same section of a track. Usually one of the two offerings is correct but also adding an incorrect homophone or pseudocognate, but sometimes both ideas are erroneous. All examples are given in Table 34.

As with the examples of metathesis, they come from a range of participants, although we can see here that they are clustered towards the end of the data collection period. Such a distribution might reflect a growing vocabulary: the very potential for a participant to offer two possible translations increases with vocabulary size. Although there is no evidence from the dataset that allows us to infer definitively that these errors stem from the auditory nature of the input, it seems reasonable to assume that they have arisen due to time pressure and the necessity of rapid processing without opportunity for reflection. Given what we have seen that listening performance appears to be driven by acoustic associations, if the sounds associate equally strongly with two candidates, it seems understandable that in some cases, both are given.

Table 34: All examples within the dataset of a double translation

Location	Input	Correct translation	Mistranslation	Explanation
T2 AS7	<i>Mon professeur</i>	My teacher	I prefer my teacher	<i>Préfère / professeur</i>
T2 St1	<i>Les profs sont gentils</i>	The teachers are kind	John is kind	<i>Gentil</i> translated both as ‘kind’ and as ‘John’
T3 E111	<i>Je ne supporte pas</i>	I can’t stand	I hate sport	<i>Supporte</i> = both ‘can’t stand’ and ‘sport’
T4 F9	<i>Je fais de l’équitation</i>	I do horseriding	I swim at the lake	<i>L’équitation</i> has been confused for <i>natation</i> (swimming) and missegmented to make ‘lake’
T8 D8	<i>Avec mes copines</i>	With my friends	Camping with my friends	<i>Copines</i> = both ‘friends’ and ‘camping’
T8 E3	<i>Ma grande sœur</i>	My big sister	The grocery store with my grandad	<i>Grande sœur</i> = grandad (grand sir?) and sounds like ‘grocer’
T10 E111	<i>Ma matière préférée</i>	My favourite subject	Maths is my favourite subject	<i>Matière</i> = ‘subject’ and missegmented for ‘maths’
T10 F9	<i>Je fais de l’équitation</i>	I do horseriding	I swim at the lake	[see above]
T12 C17	<i>Que manges-tu ?</i>	What do you eat?	I eat mangetout	<i>Mange</i> = ‘eat’, also missegmented with <i>tu</i> added
T12 D8	<i>La musique rap</i>	Rap music	Rap music and hard rock	<i>Rap</i> = both ‘rap’ and ‘rock’
T12 E26	<i>Mon portable</i>	My mobile phone	My portable phone	<i>Portable</i> = both ‘mobile phone’ and ‘portable’
T13 E56	<i>Au bord de la mer</i>	At the seaside	By the sea with my mum	<i>Mer</i> = ‘sea’ and ‘mum’ (<i>mère</i>)
T16 D8	<i>Je fais du vélo</i>	I do biking	I’m going to ride my bike	<i>je fais</i> = ‘I do’ as well as ‘I’m going to ride’
T16 E3	<i>Ma sœur préférée</i>	My favourite sister	I prefer my professor	<i>Préfère</i> as above
T16 C41 tr	<i>On peut aller</i>	We can go	In a bit we can go	<i>On peut</i> = we can; <i>un peu</i> = a bit
T16 C41 tr	<i>Bouteille d’eau</i>	Bottle of water	Buttered potato	<i>Bouteille d’eau</i> = potato, with <i>bouteille</i> also being mistranslated as ‘butter’
T16 C41 tr	<i>Les choses sucrées</i>	Sugary things	Hot sugary things	<i>Choses</i> = things, also being mistaken for <i>chaud</i> (‘hot’)
T17 C48	<i>Partager ma chambre</i>	to share my room	share part of my room	<i>Partager</i> = to share. Also missegmented and mistranslated for ‘part’
T18 St3AP	<i>On est allé</i>	We went	We went to Italy (x2)	<i>On est allé</i> sounds like <i>en Italie</i> in Italy
T19 St351A	<i>Dans ma chambre</i>	In my room	A charger in my room	<i>Chambre</i> for both ‘room’ and ‘charger’
T20 St3M42A	<i>Au bord de la mer</i>	At the seaside	Skateboard by the seaside	<i>Au bord de</i> for both ‘at the side of’ and ‘skateboard’

4.3.4. Active knowledge is no guarantee of passive recognition

At Time 20, the phrase *alors William*³⁵ /alɔʁ wiliam/ (so, William) was translated by 38 out of 54 participants as *hello William*, even though it seems reasonable to assume that if the same 54 participants were asked ‘how do you say hello in French’, they would answer ‘*bonjour*’. In this case, then, 38 participants have been thrown by a word they do not know – *alors* – made a guess at its meaning given its phonological proximity to the English ‘hello’.

In other words, active knowledge of words does not appear to guarantee recognition within the demands of listening – and in the case cited above, its phonology overrides correct meaning for more than half of the participants. In fact, in many cases, a participant would correctly translate a word in one track, only to use the same translation for a different word in the next. An example of this occurs in *j’aime regarder les films français*³⁶ (*I like to watch French films*) followed by *mais ce soir il y a beaucoup de feuilletons américains* (*but tonight there are a lot of American soap operas*). This is illustrated in Figure 33, below.

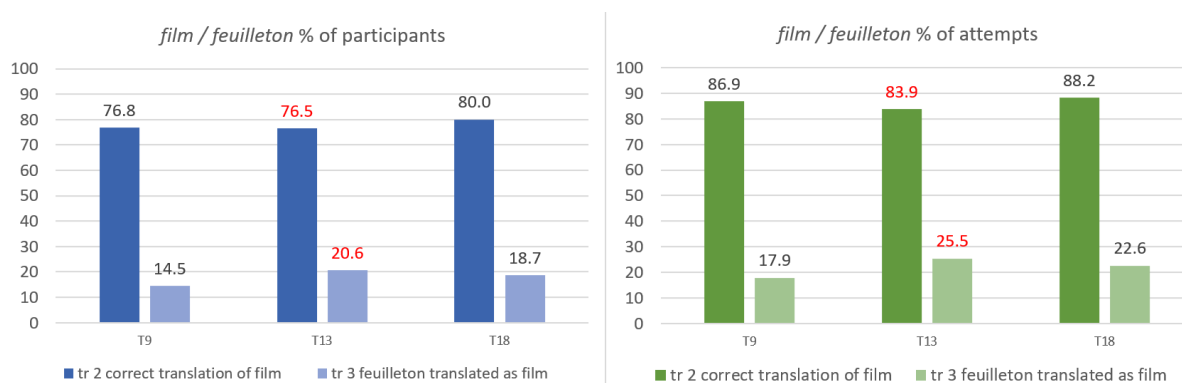


Figure 33: Correct translation instantly followed by incorrect translation

It is no surprise that in track 2, most participants correctly translated *films*; yet many of them still translated *feuilletons* as *films* in the subsequent track, and that with time,

³⁵ in test St342A, T20

³⁶ tracks 2 and 3 of test D25, (T9 T13 T18)

progress does not appear to be made (All participants who wrote *feuilleton* had also written *film* at all three time points.) This phenomenon exemplifies the demands of working at speed combined with the dynamism of the listening experience – where it is likely that the *feuilleton* / *film* error is in part driven by phonology, and *alors* / *hello* by an interaction of phonology and context. There are further examples within the dataset, such as *tous* being incorrectly translated as ‘two’³⁷, or *ville* being correctly translated as ‘town’ but *village* is then also translated as *town*³⁸, driven by false friends and lexical sets, and the overlooking of *avec* in the phrase *avec ma mère* (‘with my mother’) resulting in a translation of ‘sea’ (*mer*) which demonstrates the content/function word relationship (see section 4.5.1) as well as a possible ‘make do’ strategy.

4.3.5. Changes over time in phonological errors

Sections 4.3.1 to 4.3.4 have addressed a range of specific errors made in the paused translation tests and begun to explore what these errors reveal about participants’ experience of listening. We now move on to consider change over time in the errors.

Over the period of data collection, which was three academic years, it might be reasonable to expect an increased accuracy of translation, and with it, a drop in errors. Recall that in section 4.1.2 I have already ruled out significant loss of learning due to the disturbances created by the pandemic. While within the dataset as a whole we have seen that consonants drive participants’ (mis)understanding, and vowels are overlooked, it might be reasonable to expect that participants become better at processing vowel sounds, and that minimal pairs might become more reliably translated. Finally, as

³⁷ Test C48 T14

³⁸ Test G32 T6

participants mature, one might expect their translations to become more sensible and less random – in other words, that sound no longer trumps sense.

Number of vocabulary errors recorded

In fact, the number of vocabulary errors did not diminish much, as can be seen from Figure 34. There are a number of possible reasons for this, the most likely reflecting the fact that the difficulty level of the texts kept pace with the proficiency of the participants.

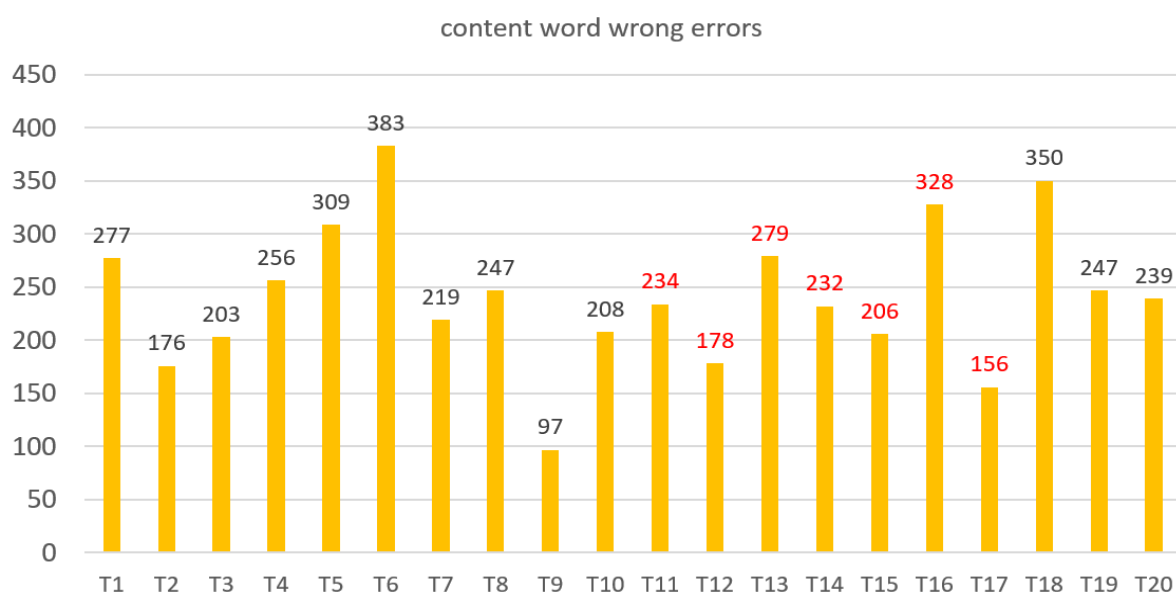


Figure 34: Number of 'content word wrong' errors recorded at each time point

In order to interrogate whether the nature of errors changed over time demands reference to the tests that were repeated over several time points.

Change in nature of phonological errors

There are various figures above which detail the change – or lack of it – over time in phonological errors. The perception of voiced and unvoiced pairs did not appear to

improve: see, for example, a very small, but unchanging number of participants who perceived *parce que*³⁹ as ‘basket’ (section 4.3.2).

We can also see a lack of development of detailed listening necessary to perceive the difference between words with similar consonants but different vowels. Section 4.3.3, which illustrates the confusion over *copain*, *campagne* and *camping*, also indicates no evidence of an increased ability to perceive different vowel sounds over the data collection period, and indeed in the case of *dessins animés* being metathesised to become *cinéma* there is even a small increase (Table 33, section 4.3.3). The pattern appears to be highly dependent on specific words, however: the mistranslations and missegmentations initially seen whereby *portable* became ‘tablet’ have dropped, as will be seen in section 4.4.2, this might be for reasons of frequency of this particular vocabulary item.

I also hypothesised that participants’ guesses at unfamiliar words might become more sensible with increased maturity, and this can be seen in Figure 29 (section 4.3.1), where the difficulty of segmenting *c’est utile* had arisen, and with time, more contextually feasible attempts at *utile* became apparent. Yet, once again, this was not a consistent picture, as can be seen in Table 30, where, even towards the end of the data collection period, *ne m’intéresse pas* was still being translated as *spa* in a passage about television. In the same test, *mais c’est beaucoup* continued to be problematic: potentially the frequency of the French phrase *merci beaucoup* for many participants overrode any common sense of the appropriacy of the phrase within the dialogue they were hearing.

³⁹ T3 1x; T10 3x; T14 1x; T18 4x

4.3.6. Summary of lexico-phonological findings

- The potential for a short stream of speech to appear to contain an English word or a French-English cognate might be a trigger for lexical or segmentation error. This finding could not have arisen if I had used paused transcription rather than paused translation, and offers important new insights into the experience of L2 beginner listening.
- As a rule, sound overrides sense: participants in the present study tend to force English words they believe they have heard into any kind of translation, regardless of how nonsensical this might be.
- An increase in the raw number of segmentation errors and vocabulary over time appears to be due at least in part to an increase in attempts at translation.
- As the time period progresses, the chances of a lexical or segmentation error stemming from French vocabulary interference (as opposed to English) increases, but remains low.

4.4. Textually driven vocabulary errors

This section discusses the vocabulary-related findings which are attributed to difficulties in understanding the content of the French, rather than the auditory input. In other words, one might assume that the difficulties discussed here could equally arise if participants had been given written texts to translate, rather than listening passages. Coding and analysis of the dataset resulted in three sub categories: false friends that are textually driven, errors arising from lexical sets, and teaching dependent errors. Again I will provide an overview of the nature of errors, then move onto a discussion of their development over time, and contextualise the findings within the wider literature relating to the fields.

4.4.1. Textual false friends

The textual false friend is a word that in writing means one thing in the L1 and another in the L2, for example *le skate* ('skateboarding'). False friends are similar to, yet distinct from the category of pseudocognates discussed in section 4.3.1. In the present study I have taken pseudocognates to mean words that are homophonous – this can be French to English – for example *alors* / hello (section 4.3.4) or *bouteille d'eau* / potato (section 4.3.1) – or French to French, such as *amuser* / *musée* (section 4.3.1). Within the dataset, my initial 'false friend' code comprised both phonological and written false friends, and when separated, 1057 were phonological and 497 were textual. Given the relatively small number occurring within the dataset, I will discuss them generally, plus their changes over time, in a single section.

There was a small number of false friends whose faulty translations would have caused identical problems had the task been a written one. Key among here was the phrase *je reste*, meaning 'I stay' rather than 'I rest', and *cadette* meaning 'youngest' (in the context of youngest child), rather than 'cadet'. As we have seen in previous cases, incidences of mistranslation increase rather than decrease for both, as set out in Figure 35 and Figure 36, below, again suggesting an increase in willingness to try (and risk failure), which could be a precursor to successful attempts.

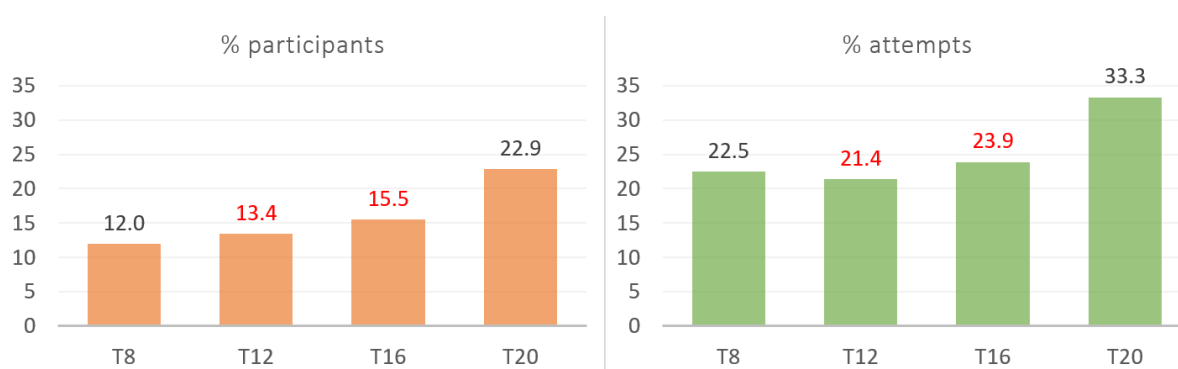


Figure 35: Change over time in translating *je reste* as 'I rest'

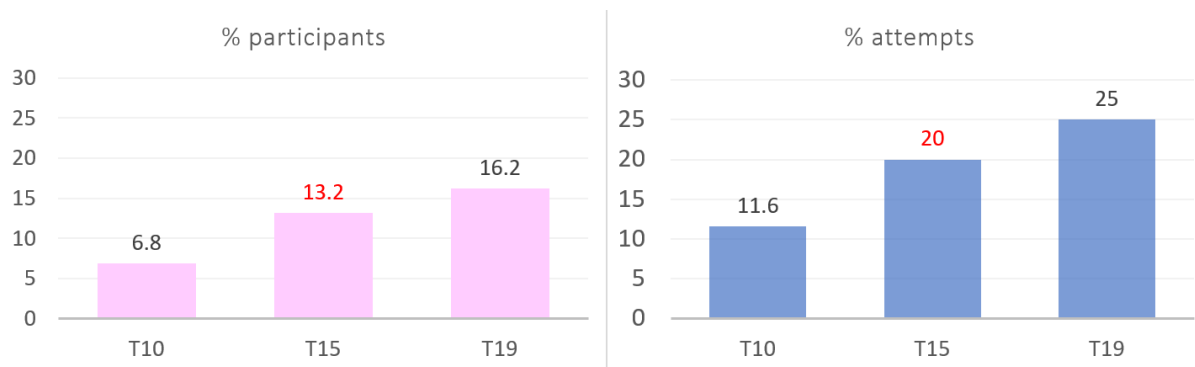


Figure 36: Change over time translating cadette as 'cadet'

Just like the phonological errors discussed at length in section 4.3, many textual false friends do not stand in isolation with the myriad other elements contributing to a translated sentence, and the choices made by my participants in the translation of *Les chips? J'adore les chips*⁴⁰ /le ʃɪps zadɔʁ le ʃɪps/ ('Crisps? I love crisps!') illustrates this well (Figure 37). For example, although in writing it is reasonable to assume many participants would translate *chips* as 'chips', it is unlikely that in a written modality participants would opt for translations such as 'sheep' 'ships', 'sheet'.

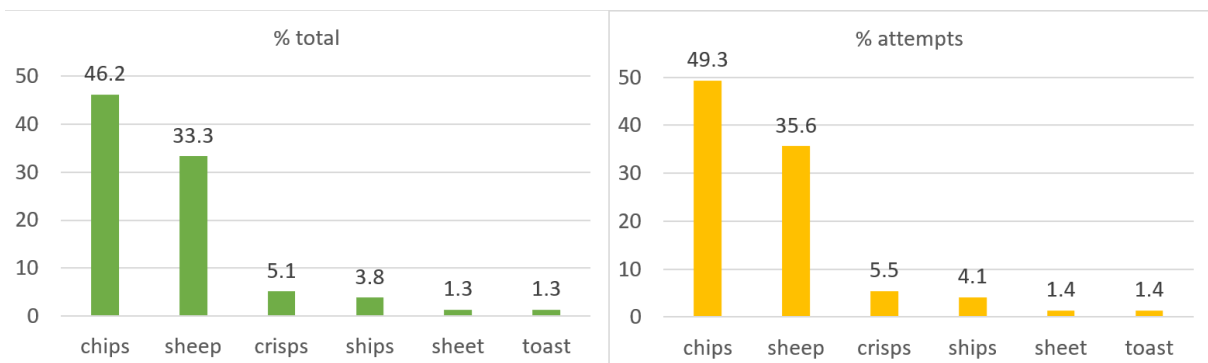


Figure 37: Phonological and textual errors interact with false friends

⁴⁰ Test F28 T5

4.4.2. Lexical sets

Two lexical sets caused many translation problems among the participants – family members, and time phrases. Nine per cent of all the ‘word wrong’ codes were assigned to these lexical set problems, of which 61% (286/468) were family members, and 39% (182/468) were time phrases. Again, these errors are *predominantly* textual rather than phonological in origin, but one cannot argue that they are *entirely* textual: evidence from the present dataset again points to the fact that text and process are interdependent. For example we see 76 instances out of 286 (26.6%) where *frère* (‘brother’) was translated as ‘friend’. The same argument can be made in the confusion between *toujours* /tuzuz/ (‘always’) and *tous les jours* /tulezuz/ (‘every day’), which also needs to be considered within the context of the discussion below on content and function words (section 4.5.1), and the lack of salience of *les* within the chunk. As discussed earlier, the methodology of this chapter did not allow for me to tease out whether this was a textual or phonological problem, although the stimulated recalls for case studies in the third findings chapter will allow some opportunity for this.

To address development over time, Table 35 shows the number of both lexical sets coded to each of the tests which were repeated. We can see that the ‘family and friends’ lexical set appears to remain constant or improve over time, whereas the ‘time phrase’ lexical set appears to remain constant or deteriorate over time. One can see, then, that my participants made progress in different lexical sets at different paces; there will be more discussion of this in Chapter 10.

Table 35: Total number of references of the lexical set codes

Test E3	T5	T8	T11	T16
total # family friends codes	26	33	17	17
total # wrong time phrase codes	9	4	14	15
Test D3	T10	T15	T19	
total # family friends codes	11	15	12	
Test E56	T6	T13	T19	
total # family friends codes	11	12	10	
Test C48	T11	T14	T17	
total # family friends codes	17	12	10	
Test E111	T3	T10	T14	T18
total # wrong time phrase codes	1	9	11	11
Test F9	T4	T7	T10	
total # wrong time phrase codes	9	7	18	
Test E26	T4	T7	T12	T15
total # wrong time phrase codes	3	2	1	2

4.4.3. Teaching-dependent errors

This section discusses errors which appear to derive from an overapplication of what is covered in the classroom, and again appear to be textual in origin rather than phonological. The phrase ‘I like’ is a significant focus with the beginner French syllabus in England, and it became clear early in coding on that within the ‘wrong word’ vocabulary category, a subcode of ‘*I like*⁴¹ for a different verb’ was needed; this code arises 748 times within the dataset (accounting for 15% of all the ‘content word wrong’ codes), in 52 of the 60 tests. In the remaining eight, ‘I like’ is the dominant verb form anyway. Again, the complex dynamic system of the individual listening experience arises: in 33 cases within the dataset, phonological interaction was apparent in that *j’ai* /ʒɛ/ (I have) was translated as *I like*. The overuse of ‘I like’ might also be a particular

⁴¹ In French: *j’aime* /ʒɛm/

problem in listening (as opposed to reading) because of the necessity to process the input rapidly without the opportunity for the listener to go back and check or reflect, combined with a learned automaticity to default to ‘I like’.

Figure 38, below, sets out the appearance of an incorrect ‘I like’ within the dataset at each time point. Recall as usual that all the errors are input-dependent: the particularly low number of codes at Time 6, for example, is because the actual input in two of the tests revolved around the phrase *j’aime* – in other words, where participants wrote ‘I like’, it was actually correct. The third test at Time 6 was hard enough that only few participants even attempted it – so did not even manage to submit an ‘incorrect I like’.

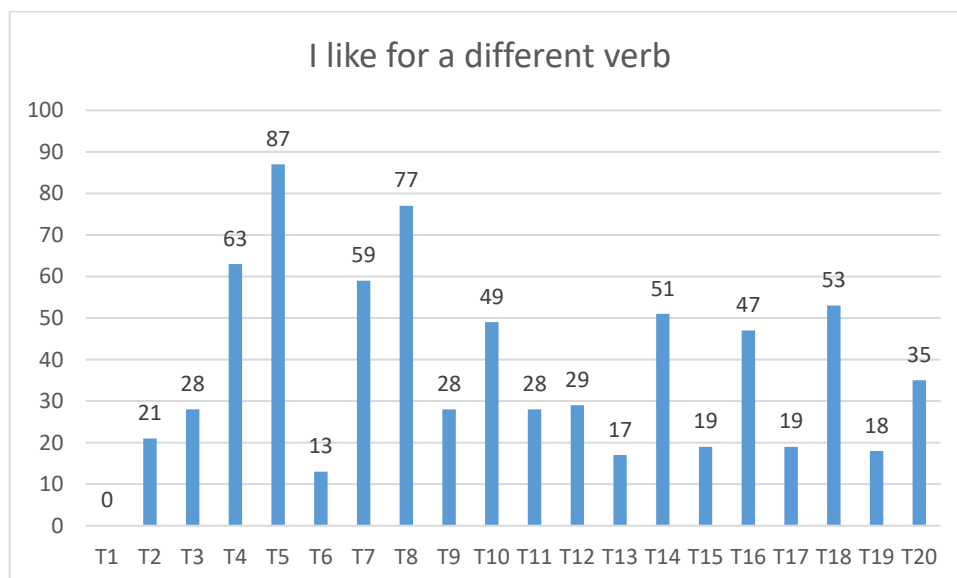


Figure 38: The appearance of an incorrect I like at each time point

When looking at change over time in individual tests, I took the four repeated tests which contained the most incorrect *I likes*: the numbers are set out below in Table 36.

Table 36: Incorrect 'I like's in repeated tests

E3	T5	T8	T11	T16
# wrong 'I like's	41	16	13	9

E111	T3	T10	T14	T18
# wrong 'I like's	26	37	30	25

E26	T4	T7	T12	T15
# wrong 'I like's	36	36	12	14

D8	T8	T12	T16	T20
# wrong 'I like's	32	19	23	19

This suggests progress over time, whereby participants are tending not to default to an incorrect 'I like' as their learning develops. Furthermore, when these data are visualised against the average engagement data for each test at each timepoint (Figure 39), we see that there is an increase in effort at the same time as the instances of 'I like's drops: genuine progress is being made. Discussion of this issue within the context of the literature will be presented in Chapter 10.



Figure 39: Default to 'I like' cross-referenced with engagement

4.4.4. Summary of textually driven vocabulary errors

Above we saw findings relating to textually driven vocabulary errors, which addressed false friends, lexical sets and teaching dependent errors. To summarise:

- False friends presented a challenge throughout the data collection period, often exacerbated by the aural modality.
- Lexical sets were also problematic, although the data made some suggestion that the nature (content versus function words) or type of lexical set might influence how problematic.
- Transfer of training was apparent in the overapplication of ‘I like’ for a variety of other verbs, and this diminished with time.

Each one of these types of error illustrates a situation in which a variety of different words in the participant’s lexicon are competing for potential answers.

4.5. Grammatical errors

The following section on paused translation errors of a syntactic nature again straddles both the textual and the phonological. I will discuss the roles of content and function words (while accepting that my choice to position this under a ‘grammar’ subheading is subjective), and the recognition of morphemes. Given that these are relatively short sections, I will incorporate general observations plus change over time into two single sections.

4.5.1. Content and function words

The role and relevance of content versus function words is a source of curiosity among scholars (Field, 2008c; Graham & Santos, 2013; Lange, 2018; Patterson, 2021, 2022), and indeed ‘function word wrong’ was the very first code identified when beginning the

error analysis in the current dataset. Table 37 reminds us of the relevant sections of the codebook summary given in Table 26.

Table 37: Codebook summary words missing vs words wrong

Code / sub code	Number of files	Number of references
word missing	60	3401
content word missing	60	1735
function word missing	46	1666
word wrong	60	5216
content word wrong	60	4991
function word wrong	42	225

Within the dataset, words were much more likely to be wrong (5279) than missing (3412). This is logical in that, where a participant did not know a word, there are two options open to them – guess, or leave it out. However, entries for ‘word wrong’ are not confined to guesses, but also to genuine errors. In other words, there are two possible reasons for a ‘word wrong’ entry, but only one for a ‘word missing’.

Yet, when we compare the relative numbers for function and content words missing versus wrong, something intriguing emerges from the data. For missing words, the number of content and function words is fairly close. But for wrong words, we can see content words dominating by a factor of 20. When viewed the other way, content words are much more likely to be wrong, but function words are much more likely to be missing. This is set out in in Table 38.

Table 38: Missing and wrong content and function words

	Missing		Wrong		Total	
	Raw number	%	Raw number	%	Raw number	%
Content word	1735	20	4991	58	6726	78
Function word	1666	19	225	3	1891	22
TOTAL	3401	39	5216	61	8621	100

The findings deserve more nuance, however: we see distinct interactions with other elements of listening when examining function and content words. Such interactions are

characteristic of the broader findings of the research question and again point to the complex dynamic system at play in the listener's experience. That participants' understanding of content words drives their inclusion of function words is evident, for example, in test D3, *mon mari Jacques* ('my husband Jacques') where participants overlooked the relationship between the word *mon* ('my'), preferring to focus on *mari* ('husband'), often translating it as the woman's name Marie. In other words, they ignored both the function word 'my' and indeed the morphemic information of masculinity it carries within it, even with the hint of *Jacques* in the sentence.

The above example demonstrated participants' overlooking a function word, but there are also examples in the data of English functors appearing in participants' translations where they did not exist in the French original, in order for their translation to make sense. Both models are driven by a domination of the content word over the function word. Even when the participants are approaching an intermediate proficiency level, confounds interfere. This can be seen in test E56, where, for many participants *mer* ('sea') in *au bord de la mer* ('at the seaside') is understood as the homophonous *mère* ('mother'). The choice of 'mother' necessitates a strategic inclusion of 'with' although it is not present in the input. Confounds are also evident in translation of *j'habite sur une ferme à la campagne*⁴² ('I live on a farm in the country'). Again, participants' translations of *campagne* ('country') as *copains* ('friends') leads to a necessary overriding of the function word *à* ('in', 'at') and in its place a strategic yet incorrect inclusion of 'with' to complement the 'friends', 'companion' or 'partner'.

⁴² track 1, test D47, T11, T17

From the present dataset we can look at whether the number of missing or incorrect function words reduces over time, by looking at the presence of functors over various repeated tests. When examining progress over time, Test E3⁴³ had the largest recorded number of missing function words, which in raw data decreased over time. Figure 40, below, demonstrates the correct translation (i.e., not missing) of the functors *mais* ('but') and *quand* ('when'). Other repeated tests show a similar pattern: that is, correct recognition of functors improves over time, when there are no other distractors.

Saliency of certain sounds plays a role, too. Which sounds become salient to the participants and when, can sometimes be unpredictable. This is best illustrated with the function word *seulement* /sœlmã/ ('only')⁴⁴. Between Time 8 and Time 11, a new high-profile student joined the cohort, whose name sounded very much like /sœlmã/; and at Time 11, there were six translations of *seulement* as his name. Before his arrival, no student attempted a name in this translation, and by Time 16 only one participant repeated the error. This level of unpredictability and specificity to the context of the learning is a key element of a complex dynamic system (Larsen-Freeman, 2017).

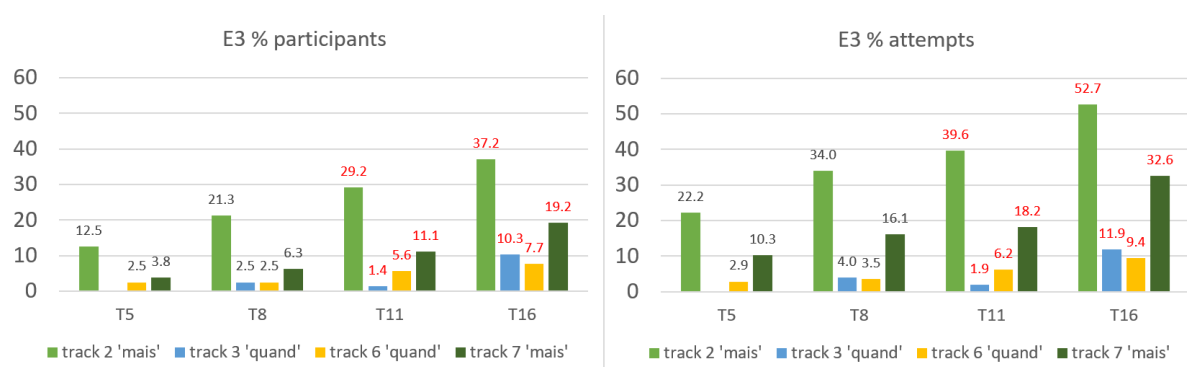


Figure 40: Change in function word appearance in test E3

⁴³ T5, T8, T11, T16.

⁴⁴ Track 7 *mais quelquefois il joue seulement avec ses amis* ('but sometimes he plays only with his friends') in test E3 (T5, T8, T11, T16)

4.5.2. Difficulty in recognising morphological features

A summary of the grammatical issues coded into the dataset is given in Table 39, below. It is important to reiterate the two key principles of this findings chapter. Firstly that participants' errors interact and overlap, and as a result, many participants' translations are coded to more than one code (for example here, 'wrong tense' is often also coded to 'hasn't processed verb ending'). Secondly, given the ecologically valid nature of the input (ie it was not carefully designed for laboratory-based tests to probe certain characteristics of learning), the numbers cited in Table 39 are dependent on the input – for example, the 'wrong tense' code arises much more frequently later in the project as, because at the beginning, all input is in the present tense. Similarly, a code such as 'didn't detect plurals' simply cannot apply to input which is only in the singular.

Table 39: Codebook breakdown of grammatical errors

	Number of files	Number of references
Grammar	58	2346
Didn't detect plurals	25	625
Hasn't processed verb endings	37	858
Incorrect negative added	24	73
Incorrect plural given	14	74
Missed negative	14	86
Wrong tense	32	630

As we have seen throughout this chapter, words, errors, and misunderstandings do not exist in isolation, but regularly interact with each other – and the references coded to grammatical errors are no different. In fact, they most commonly interact with the challenges of content and function words. There is a neat example of this is one participant's translation of *mais le prix est trop cher*⁴⁵ where *prix* ('price') is

⁴⁵ (track 6 test C11 T17),

mistranslated as the homophone *pris* ('taken'). This error demonstrates a failure to apply grammatical knowledge: the definite article *le* ('the'), which should identify *prix* as a noun is ignored. All the precursors that would be necessary if *pris* were a past participle, such as a pronoun and auxiliary verb, are absent from the sentence. In short, no grammatical error exists in isolation.

An unexpected finding regarding the influence of both the input and the learning context arose when examining the 630 codes to 'wrong tense', in that 219 of them (34.8%) can be traced back to tests on the theme of holidays or travel abroad. I attribute this to transfer of training both at the level of L2 learning, and more generally (section 4.4.3). It is a classic approach in language learning in the English school system to use a topic of holidays to teach and revise both past and future tenses (C. Bell, McLachlan, & Ramage, 2016), but also engrained in many pupils as a standard writing activity at the beginning of a new school year to write about past holidays.

It is most appropriate to compare the repeated tests over time in order to probe the nature of grammatical errors in listening, and test D8 uncovers some phenomena on noticing plurals, which are detailed in Figure 41.

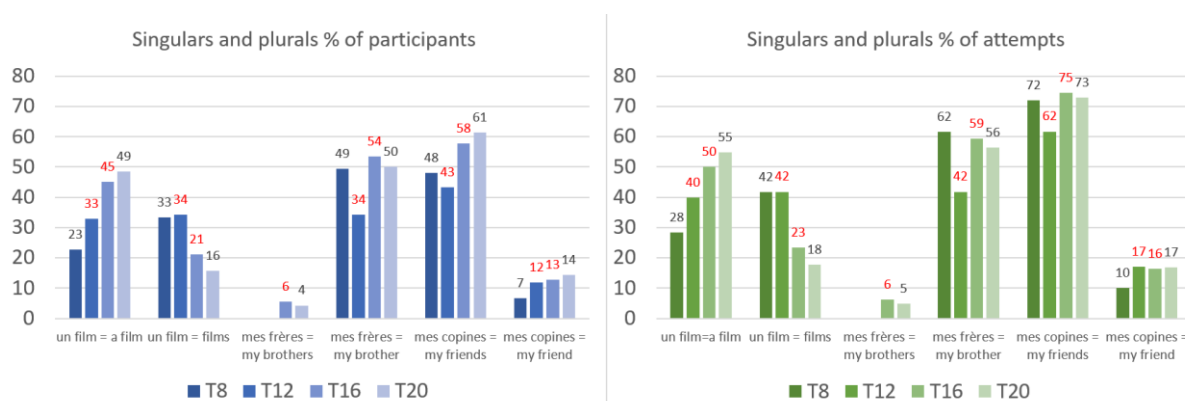


Figure 41: Grammatical errors of plurality in test D8

Je regarde un film d'épouvante avec mes frères ('I watch a horror film with my brothers') is a sentence which tests plurality twice. Translating *un film* as 'films' demonstrates a

failure to recognise the word *un* ('a') as well as a failure to grasp that this article denotes singularity and would not be correct in the case of plural 'films' (where *des* would be used instead). It is notable that although more participants note the singular *un film* correctly over time, the number who continue to mistranslate *un film* in the plural does not fall until Time 16. In other words, the percentage of correct translation does not change in direct proportion to the percentage of incorrect.

The difference in translation of *mes* ('my' [plural]) between track 3 (*avec mes frères*, cited above) and track 7 (*je fais du vélo avec mes copines* ('I go cycling with my girlfriends')) suggests that the driver for correct plural use is more founded in context or participants' life experience than it is in noticing the difference between *mon*, *ma*, and *mes*: I hypothesise that participants' choice of plural is based on an unconscious assumption that a speaker is more likely to have multiple friends than multiple brothers.

Codings to 'wrong tense' are scattergun, with the exception of test D25, which featured two tracks in the near future tense. These flummoxed most participants at all three time points. However, it also highlighted a general failure to recognise verb endings, and again we see two elements working together: whether the input is *je regarde* ('I watch'), *je vais regarder* ('I'm going to watch'), *j'ai regardé* ('I have watched') or *j'aime regarder* ('I like watching'), the participant's focus is on the word *watch*. Even at the end of the data collection period, the correct tense or grammatical function is largely overlooked. In other words, auxiliaries *je vais* ('I'm going') or *j'aime* ('I like') or *j'ai* ('I have') followed by the correct form of the verb are not being differentiated in these tasks from a conjugated verb in the present tense. The overcorrection in favour of *I like* as discussed in 'teaching-dependent errors' in section 4.4.3 adds another layer of dynamism to the grammatical coding here. The jumble of confused answers and lack of discernible progress within the cohort is demonstrated in Figure 42.

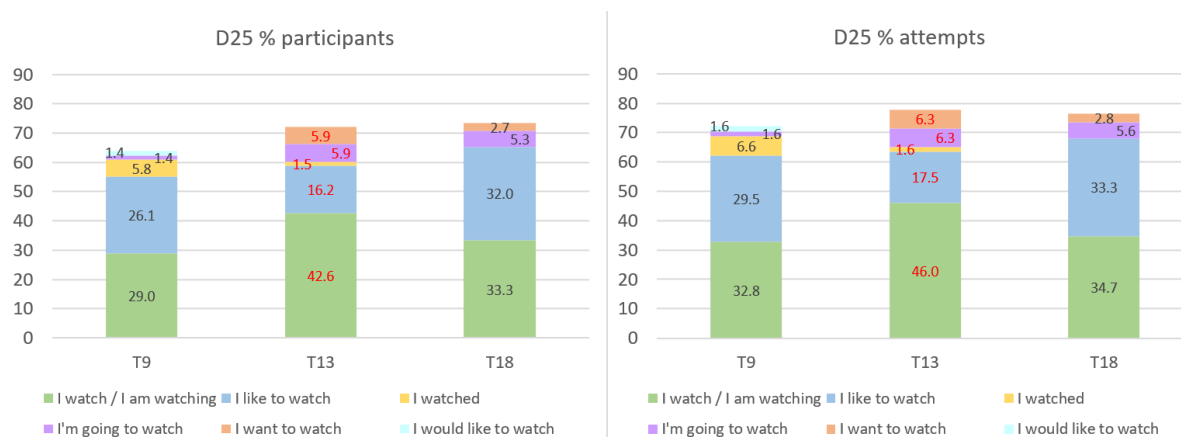


Figure 42: Translations of 'j' aime regarder' in test D25

4.5.3. Summary of grammatical errors

Above, the perception of content and function words were discussed as a principle grammatical error, followed by difficulty recognising morphological features. To summarise:

- Where a participant encounters a difficult word, content words are more likely to be wrong, yet function words are more likely to be missing.
- There is some evidence that participants become better at recognising function words over time.
- World knowledge, general knowledge or common sense might be a greater driver of participants' choice of morpheme than input.
- There is no evidence in the dataset of progress in correct recognition of verb endings to indicate tense.

4.6. Overarching conclusion of chapter

The findings from this chapter above can be analysed at two levels. There are micro-findings which have been addressed in each section – for example that participants are more likely to omit the end of a track than the beginning, or that while they progressed in

recognition of plurality, they did not improve in misapplication of plurality. At the macro level, the overarching themes to be drawn from this chapter are as follows:

- The aural modality of the listening experience is key, and the experience of processing sounds, which differs from the way in which written words would be processed, is the driving force behind most of the errors.
- For every participant, the processing of each short track, lasting only a few seconds and consisting of no more than ten words, generates a unique complex dynamic system, and contains within it its own developmental trajectory, adhering to the properties of emergentism, relativity, non-linearity and self-organisation (Larsen-Freeman & Cameron, 2008).

5. Findings 2: relationships between subsystems

This chapter reports on the quantitative findings of research question two, which is ‘how do the relationships between listening proficiency and the subsystems of language knowledge and metacognition develop over three academic years?’ It begins by examining variability of listening proficiency, operationalised with scores derived by the paused translation data and a Rasch model, and then does the same with two sets of subsystems – linguistic and metacognitive. As an aide-memoire the details of these are set out in Figure 43.

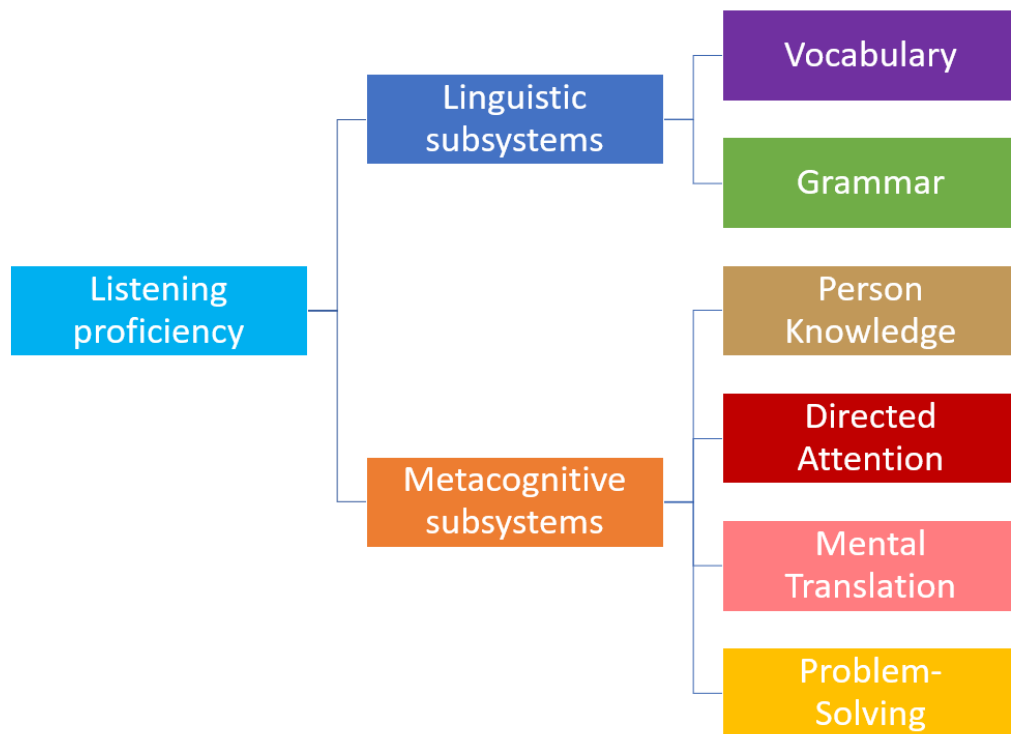


Figure 43: Reminder of the structure of the system and its subsystems

I begin by reporting on listening proficiency as a whole with a line graph including a smoothed trendline and moving min-max graph. (See methodology section 3.4.4 and 3.6.2. for details of how I reached the listening proficiency scores and section 3.6.2 for explanation as to why the smoothed trendline and moving min-max graphs are used.) I

will then examine the development of the linguistic systems of vocabulary and grammar, again using moving min-max graphs, before moving on to the relationship between the two subsystems and that of the ‘umbrella variable’ of listening proficiency by means of the CDST technique of moving correlations. This process will then be repeated with the MALQ scores of Person Knowledge, Problem-Solving, Mental Translation and Directed Attention: first individual moving min-max graphs, then moving correlations to examine the extent to which different variables play different roles at different times (See section 3.6.2 for how I reached MALQ scores,)

5.1. Listening proficiency and its variability

This section provides an overview of participants’ listening proficiency, set on a scale of 0 to 100. The score of 100 could roughly equate to a CEFR score of B2, or a GCSE grade 5 to 6. The lowest scores indicate very minimal understanding of only one or two words in a passage of 70 to 100 words or around 30 seconds.

5.1.1. Whole sample listening scores

Table 40 reports the descriptive statistics for listening proficiency (given in scores and generated by the Rasch model). Clear progress can be seen over time, with a correlation between time point and mean listening proficiency score of 0.88. The highest mean score is not at Time 20 but Time 19, and the highest individual score occurred at Time 15, by case study Theo (see chapter 7). Recall that Time 1 data collection took place after the participants had been learning French for only six weeks, but the lowest score recorded took place not at Time 1, which might be expected, but at Time 12, which was after 18 months of learning. (Note, though, that Time 12 was also the first data collection point to take place in full lockdown, in May 2020.) A widening standard deviation is also noteworthy.

Table 40: Descriptive statistics for listening proficiency

Time point	N	Minimum	Maximum	Mean	Std. Deviation
T1	77	13.7	47.4	36.4	6.2
T2	72	28.1	53.5	40.1	5.2
T3	76	27.0	56.1	40.3	6.9
T4	79	20.8	63.4	44.1	7.4
T5	77	21.0	67.6	41.9	7.5
T6	74	20.6	58.6	41.6	9.0
T7	75	17.7	72.7	47.4	8.2
T8	74	12.7	84.2	42.6	13.6
T9	69	20.7	70.4	44.6	9.7
T10	73	15.3	74.4	47.5	10.6
T11	67	26.7	70.3	48.9	10.2
T12	67	12.4	80.4	49.1	10.8
T13	66	29.2	72.2	49.3	9.5
T14	73	21.6	70.7	47.0	10.8
T15	78	28.0	100.0	51.7	11.5
T16	73	24.4	87.2	51.5	14.5
T17	71	26.7	78.1	50.9	13.3
T18	73	26.5	83.2	49.4	10.8
T19	67	23.9	98.5	52.9	14.3
T20	69	23.6	77.7	50.8	13.0

5.1.2. Variability in paused translation scores: whole sample

Initially, variability within each individual subsystem needs to be explored (for full details and philosophical underpinnings see methodology section 3.6.2.)

Mean listening proficiency scores, overlaid with a trendline (polynomial to a factor of 6), are given in Figure 44, the trendline allowing us to capture general tendencies while maintaining an element of variability (van Dijk et al., 2011). The graph illustrates what is set out in Table 40: that is, a steady increase in listening, but with fluctuations. The trendline suggests that although progress was made, it slowed significantly after Time 15, and had started to plateau after Time 17.

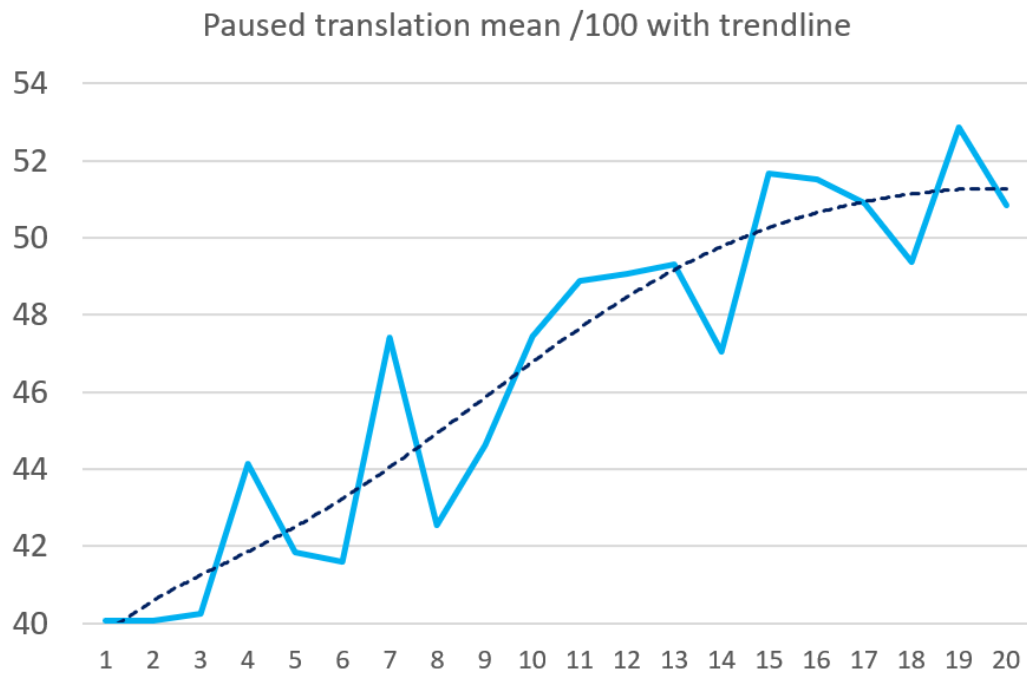


Figure 44: Mean paused translation scores plus trendline

The listening proficiency moving min-max graph is given in Figure 45, and through this we are able to explore and analyse the changes between the minimum and maximum mean scores over time (full details in methodology section 3.6.2). In this case, the distance between the moving minimum and the moving maximum widens noticeably at Time 5, suggesting a change in the nature of the group’s proficiency at that stage. The small change in moving maximum between Time 11 and Time 12, followed by the large jump between Time 12 and Time 13 is characteristic of the fluctuation that can occur before a developmental leap is made.

Mean, and moving min-max paused translation scores

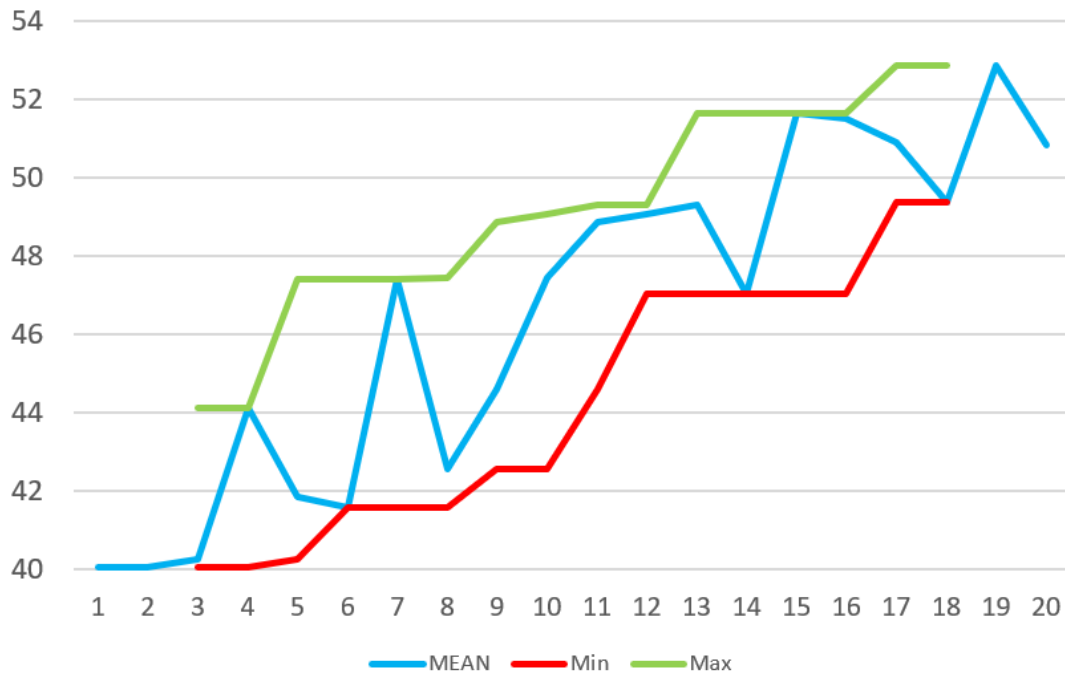


Figure 45: Moving min-max paused translation scores: mean of cohort of 80

At this juncture, then, we can look more closely at the moving min-max graph by focussing on the minimum and maximum lines and the relationship between them over time. One underpinning assumption of the present study is that development is characterised by variability: in other words, as learning takes place, performance fluctuates, and as maximal achievement is reached, performance will stabilise. In graphical terms this would mean that the bandwidth between moving minimum and moving maximum will shrink drastically. However, it is reasonable to expect some fluctuation even in a stable model. This means that the bandwidth is not likely to disappear altogether: the moving minimum and maximum are unlikely to be equal.

A wide bandwidth is equal to variability in performance and narrower bandwidth is equal to more stability or even reliability in performance.

We see a wide bandwidth at the beginning of the study, until Time 10, at which point the red ‘minimum’ line rises steeply. Given that the ‘maximum’ does not rise with the same steepness, Time 12 shows the narrowest bandwidth of the study, showing some semblance of short-lived stability, because it was at this time that we see the smallest gap between minimum and maximum scores. However, this is short-lived, and the variation picks up again, although in less marked a fashion, between Time 12 and Time 13. It steadies until Time 16, (the minimum and maximum lines are parallel), suggesting some equilibrium within the full cohort during this time, then increases. The study ends with another short period of parallel lines: another period of equilibrium with a higher minimum and maximum.

5.2. Linguistic subsystems and their variability

This section examines the individual subsystems of vocabulary knowledge, grammatical proficiency and the four individual elements of the Metacognitive Listening Questionnaire (Vandergrift et al., 2006), which are Person Knowledge, Directed Attention, Problem-Solving and Mental Translation. As above, mean scores are reported along with a smoothed trendline, and moving min-max graphs probe some of the variability within scores.

5.2.1. AuralLex scores

Recall that at each time point participants heard 48 words, 40 of which were real French words taken from the GCSE syllabus, plus eight non-words. For each correctly identified real word, they scored 50 points, and for each non-word that they said they knew, they lost 250 points. From this, an estimated vocabulary size of up to 2000 was calculated at each time point. Descriptive statistics for the AuralLex scores are in Table 41 and a graph of the mean with polynomial trendline is shown in Figure 46. As set out in methodology

3.4.5, after piloting four AuralLex tests found to be of similar difficulty, and these rotated throughout the data collection period. They are colour-coded in both Table 41 and Figure 46: in other words, the time points shaded in red were the same test, as were numbers shaded in orange, and so on. This demonstrates that the fluctuations cannot be said to be due to the specific tests, but other factors were at play.

From this we see that the highest individual score (of 1250) was achieved at Time 8, and the highest mean of 446 was achieved at Time 13, with 443 achieved at Time 4.

In other words, it appears that vocabulary size within the sample initially grew rapidly, although the widening standard deviation suggests a widening of the vocabulary proficiency range over time. Figure 47 illustrates clearly how the mean vocabulary size levelled off from Time 7 until Time 16 (throughout year 8 and into the beginning of year 9), before tailing off dramatically from Time 17 onwards. Recall that Times 11 to 14 were in lockdown, and participants returned to school at Time 15, but a second lockdown took place affecting Time 17, with participants returning to school for the second time shortly

before the Time 18 data collection. It is unclear what explains the poor performance at Times 18 to 20.

Table 41: Descriptive statistics for AuralLex scores

	N	Minimum	Maximum	Mean	Std. Deviation
vocabT1	79	0	850	266	188
vocabT2	74	0	750	280	205
vocabT3	76	0	950	246	232
vocabT4	80	0	1000	314	248
vocabT5	77	0	950	443	239
vocabT6	76	0	1050	397	230
vocabT7	74	0	1050	376	262
vocabT8	76	0	1250	402	284
vocabT9	69	0	900	409	252
vocabT10	75	0	950	389	259
vocabT11	69	0	1200	420	296
vocabT12	68	0	1150	458	303
vocabT13	70	0	1100	446	272
vocabT14	76	0	1100	391	312
vocabT15	74	0	1000	396	303
vocabT16	71	0	1150	436	301
vocabT17	72	0	1050	429	295
vocabT18	76	0	950	382	267
vocabT19	68	0	1050	376	311
vocabT20	72	0	950	350	283

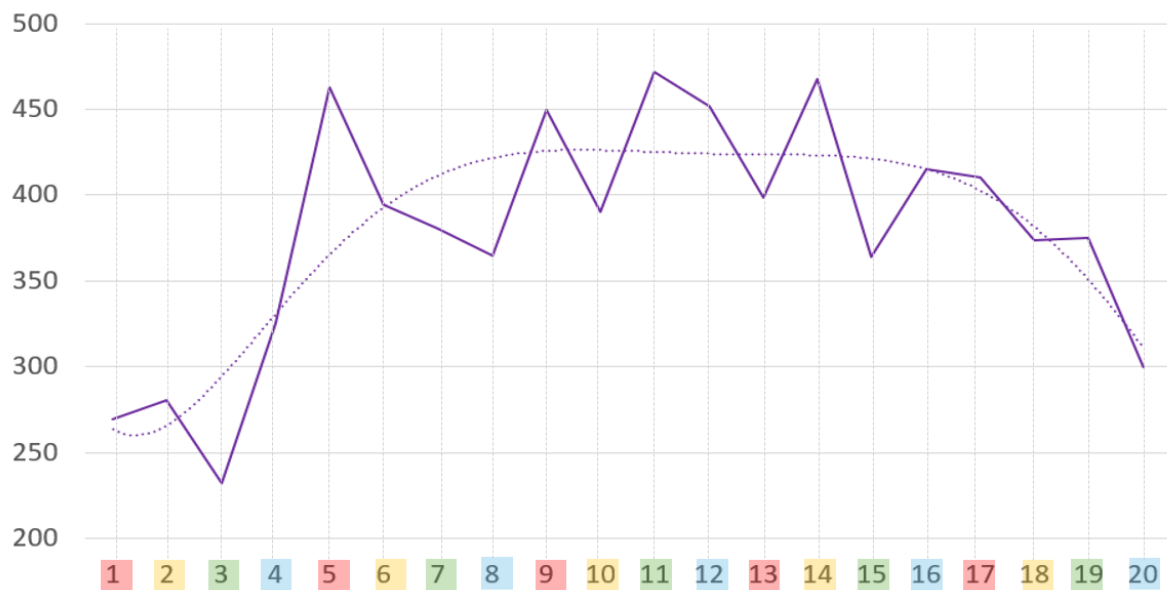


Figure 46: Whole cohort AuralLex scores, with polynomial trendline

The moving min-max scores (Figure 47) demonstrate initially a very wide bandwidth or variability, which narrows and stabilises as of Time 7, indicating again that, at a full cohort level, vocabulary recognition grew rapidly, but had peaked by the end of the participants' first year of learning. When viewed with the emphasis on the moving minimums and maximums, we can see an initially wide bandwidth representing more variability or instability of lexical performance, followed by a stability running to Time 16 (Christmas of year 9). At the end of the study, lexical performance falls dramatically, but maintaining a similar bandwidth, indicating a continued instability within the cohort.

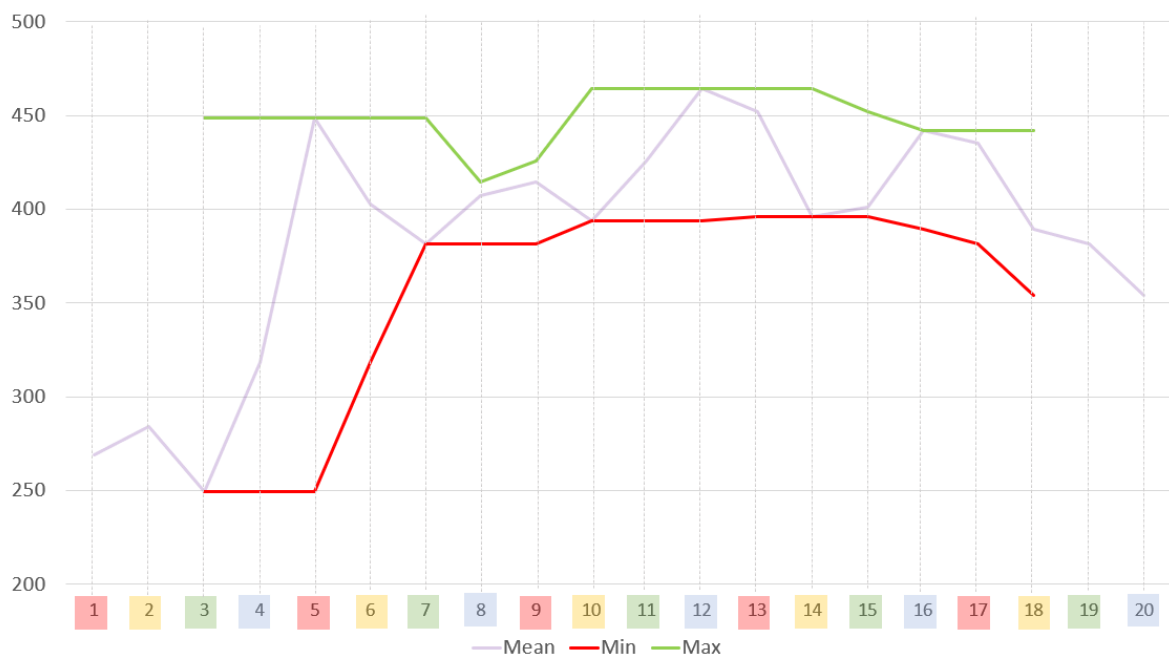


Figure 47: AuralLex moving min-max for the whole cohort

5.2.2. AuraGram scores

Grammatical knowledge was tested out of 20. I created an aural grammaticality judgement test which I called the AuraGram, whereby at each time point ten grammatical sentences and ten ungrammatical sentences were played to participants. A single point was awarded on each occasion where a participant correctly identified the grammaticality of the sentence. Full details can be found in methodology section 3.4.7. In a similar process to the AuralLex, above, after piloting six tests of equal difficulty were chosen. Descriptive statistics for the AuraGram scores are given in Table 42 and a graphical representation of the mean, with polynomial trendline, is shown in Figure 48, with the repeated tests colour-coded to demonstrate that fluctuations in score were not due to different tests.

Table 42: AuraGram descriptive statistics for whole cohort

	N	Minimum	Maximum	Mean	Std. Deviation
Gram1	77	0	12	3.623	3.052
Gram2	73	0	14	5.219	3.318
Gram3	76	0	14	4.658	3.747
Gram4	79	0	15	5.329	3.551
Gram5	76	0	14	5.789	2.891
Gram6	75	0	12	6.027	3.579
Gram7	70	0	12	5.571	2.932
Gram8	75	0	12	5.960	3.367
Gram9	68	0	11	5.632	3.376
Gram10	73	0	15	6.685	3.192
Gram11	73	0	13	6.000	3.532
Gram12	69	0	14	6.754	3.512
Gram13	72	0	13	5.875	3.224
Gram14	80	0	14	6.475	3.019
Gram15	75	0	12	6.173	3.206
Gram16	73	0	12	6.630	3.260
Gram17	76	0	12	6.526	3.231
Gram18	75	0	13	6.933	3.371
Gram19	67	0	13	5.910	3.009
Gram20	70	0	14	7.129	3.225

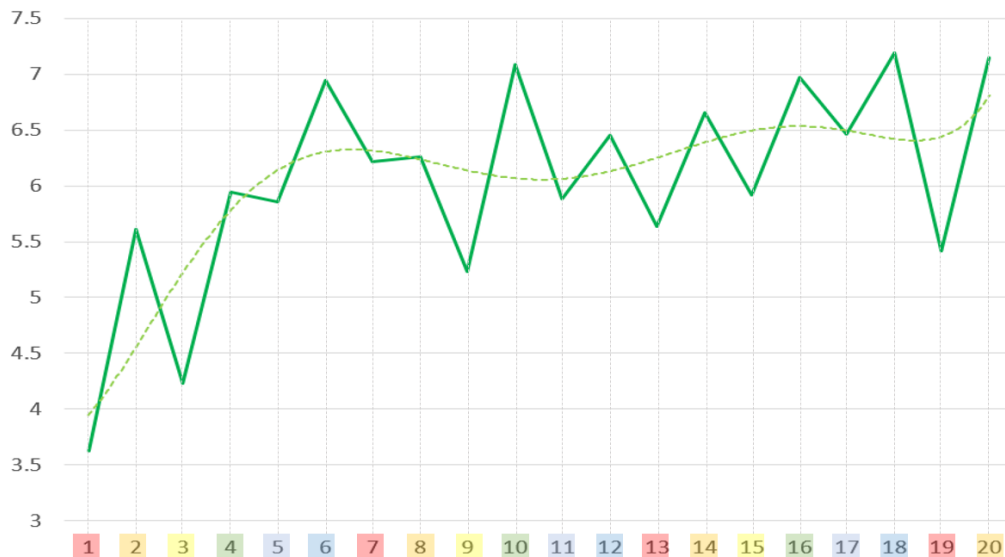


Figure 48: Mean AuraGram scores with polynomial trendline

As with the AuralLex, progress was not linear, with the highest score of 15 being achieved at Time 10, although scores of 14 were achieved at Time 12 and Time 20. Also similarly to the AuralLex, most rapid progress was made between Time 1 and Time 6 (that is, in year 7), and after this, the rate of progress falls off markedly. However, unlike with the AuralLex, the standard deviation does not widen, and the trendline shows a much more predictable pattern of progress than that of the AuralLex, with the highest mean score at Time 20.

The AuraGram moving min-max graph is shown in Figure 49.

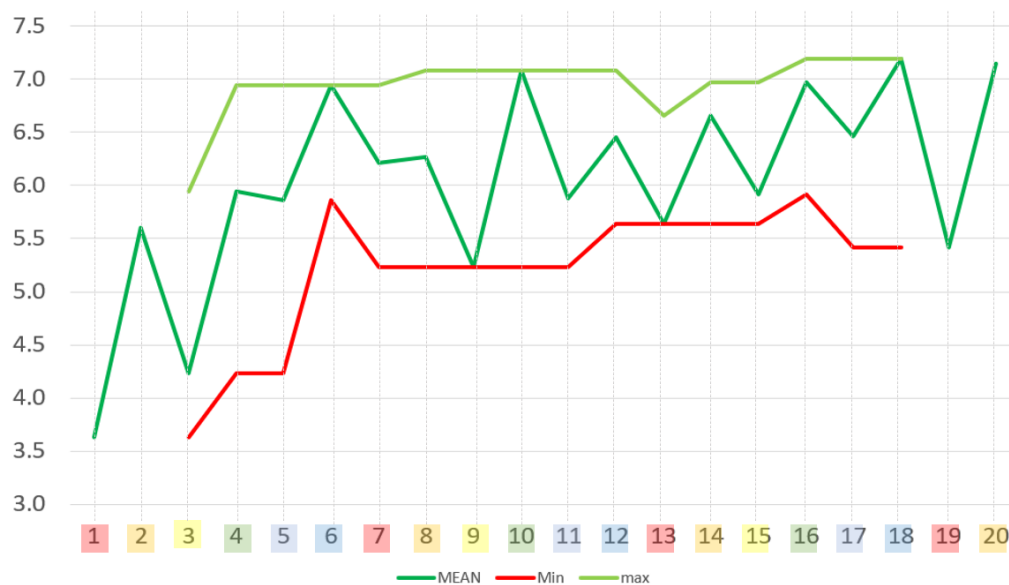


Figure 49: AuraGram moving min-max graph for full cohort

The bandwidth begins wide, indicating a great deal of variability, and then narrows, suggesting that the group might be about to reach a period of minimal change. From Time 7 to Time 11 indeed shows a smaller and stable bandwidth, during which the moving minimum and maximum scores barely change. However, there follows a brief period from Time 12 to Time 13 where the minimum and maximum scores show signs of convergence: another suggestion of a potential stable end state on the horizon. However, the bandwidth then widens and even diverges, where the increasing difference between

the moving minimums and maximums imply that the mean score is yet again in a period of flux or unpredictability.

Recognising the variability within the different individual subsystems is the first step to probing the relationships *between* listening proficiency and the subsystems, which I will now investigate.

5.2.3. Developing relationship between linguistic subsystems

So far, we have demonstrated that listening proficiency, plus the linguistic and subsystems measured, have all shown variability throughout the 20 data collection points, even when considering the mean of all 80 participants in the study.

At this juncture we aim to examine any relationships between listening and the subsystems one by one. Complex Dynamic Systems Theory states that the relationships can fall into one of three categories. Firstly, they can be supportive, whereby two elements develop in unison with one another. Secondly, they can be competitive: if this is the case, the elements will develop in alternating patterns because they compete with each other – when one is in a development stage, the system lacks capacity to allow the other to develop at the same time: instead it might stall, or even decline. Thirdly, they can be conditional: a minimal level of one subsystem is necessary before another can begin to develop: conditional relationships suggests that one subsystem's growth is the precursor to another (Verspoor & van Dijk, 2011). The relationships between subsystems are not fixed: it is a characteristic of the complexity and dynamism of a system that relationships can move between these three categories over time.

In order to best assess the status of the relationships between the subsystems and listening proficiency, all scores for all variables were normalised to a score of between 0 and 1, where 0 was the lowest score recorded during the study, and 1 the highest

(Verspoor & van Dijk, 2011). This allowed me to carry out a moving correlation analysis, with windows of five time points, to inspect the extent to which the relationships might have changed over the study period. Note that the resulting coefficients are based on small sets, with the potential for values to be overestimated and they cannot be tested for significance. This technique is entirely descriptive. For more details refer to methodology section 3.6.2.

As for the moving correlations, an entirely supportive relationship between two variables will result in a correlation of 1, and an entirely competitive relationship between two variables will result in a correlation of -1. When reporting the moving correlations graphically, all graphs will be given on a scale of 0 to 1, (or -1 to 1 if there are negative correlations to report). Where a relationship is positively conditional (ie one variable's *growth* follows that of another), an upwards curve will be visible, flattening out at a high score as the relationship becomes more supportive. And where a relationship is negatively conditional (ie one variable's growth follows the *decline* of another), the reverse will be true: a downwards curve will be visible, and flattening at a low number (-0.5 or less) as the relationship becomes more competitive.

Figure 50 illustrates listening proficiency, vocabulary and grammar development for the whole cohort, normalised to a single scale. For context, in vocabulary terms, the 0.5 line represents a vocabulary size of 600 words and 0.25 represents a vocabulary size of 300 words (as the maximum observed score across all time points was 1250 words). In grammar terms, the maximum observed score was 15/20, so the 0.45 line represents a score of 7/20.

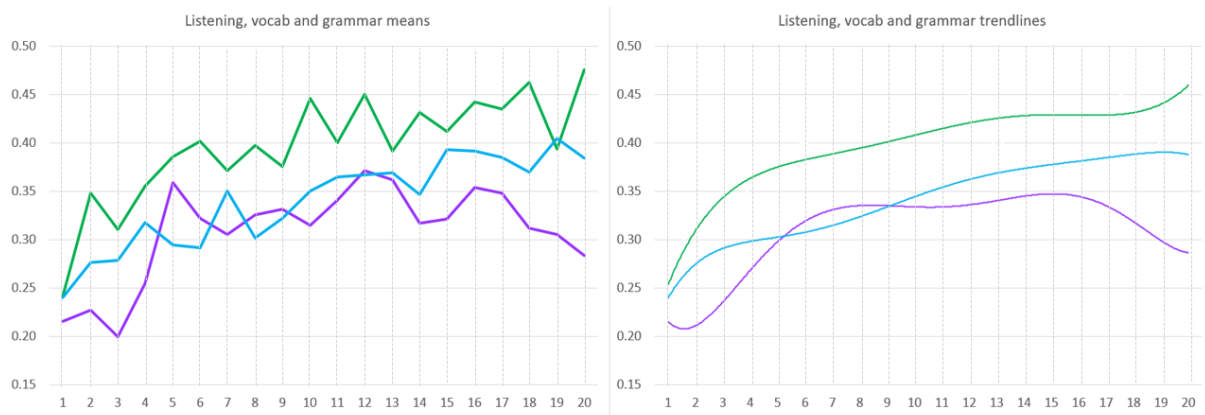


Figure 50: Listening proficiency, vocabulary and grammar on the same scale - raw scores plus trendlines

Table 43 illustrates the moving correlation between the normalised listening scores and both vocabulary and grammar. Weak correlations between 0.1 and 0.3 are marked in red, with moderate correlations of between 0.3 and 0.5 in yellow. We see from the graph that there is a moderate correlation between listening and vocabulary in almost all of the windows, rising to a coefficient of 0.476 at the Times 9-11 period. The relationship then begins to weaken, and then starts to grow again, with a second peak in the Times 17-19 window. By contrast, the relationship between the grammar scores and the listening scores, although beginning weakly, becomes negligible in the windows Times 4-6 to 7-9. It is only in the Times 16-18 window that the coefficient even becomes moderate.

The interactions between listening proficiency and vocabulary knowledge and separately for listening proficiency and grammatical knowledge are plotted in Figure 51. The plot indicates that, for the listening/vocabulary relationship, there is a consistently positive coefficient value over time of moderate strength, although that the relationship between the two variables strengthens between the Time 4-6 window and the Time 9-11 window, before weakening slightly, then becoming stronger again at the end of the study period. By contrast, the moving correlation between listening proficiency and grammatical knowledge also shows a mostly weak relationship, with a period below 0.1 (no

correlation) between Times 5-7 and Times 7-9. From Times 8-10 the relationship strengthened steadily, briefly becoming a moderate correlation in the Times 16-18 window, dropping again at the very end of the study.

Table 43: Moving correlations for linguistic subsystems for the whole cohort

	LC:vocab	LC:grammar
1-3	0.237	0.212
2-4	0.303	0.198
3-5	0.325	0.145
4-6	0.246	0.098
5-7	0.300	0.048
6-8	0.328	0.065
7-9	0.384	0.062
8-10	0.424	0.121
9-11	0.476	0.147
10-12	0.439	0.182
11-13	0.393	0.166
12-14	0.373	0.250
13-15	0.368	0.244
14-16	0.332	0.253
15-17	0.367	0.283
16-18	0.356	0.322
17-19	0.469	0.298
18-20	0.451	0.179

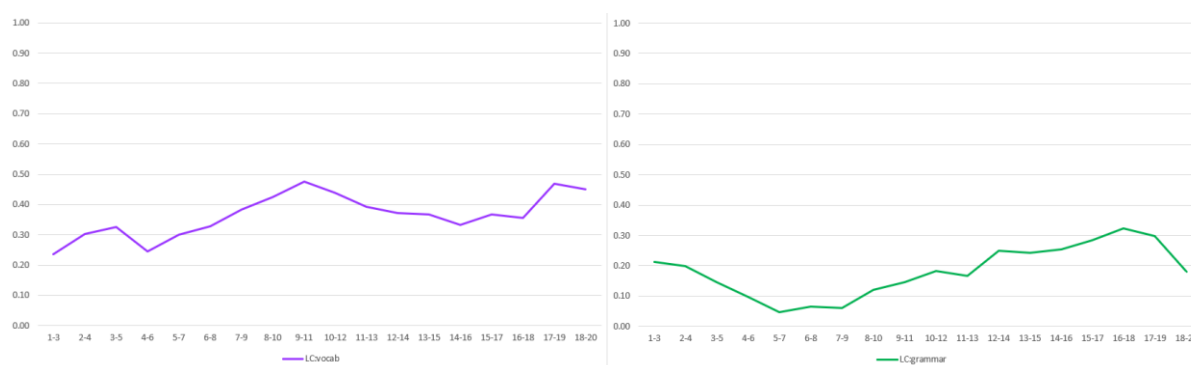


Figure 51: Whole cohort moving correlation between listening and vocabulary, and listening and grammar

Finally, Figure 52 plots both correlations on a single graph, in order to interrogate the complex dynamic system between all three variables. We see an initial divergence between correlations, hinting at possible competitive relationship between vocabulary and grammar in the beginning stages of learning. Yet from the Times 5-7 window, the

lines become parallel: in other words, the relationships begin to follow each other.

Although the correlation between grammar and listening only even becomes weak at Times 8-10, and that of vocabulary and listening is moderate, both relationships grow at a similar pace. There is a convergence of lines which begins at the Times 9-11 window and lasts until Times 14-16 and suggests another competitive relationship, although this time it is grammar which is the grower. A second period of support is then visible, and the relationships finally begin to be competitive again, where the correlation between vocabulary and listening is growing and that of grammar and listening is shrinking.

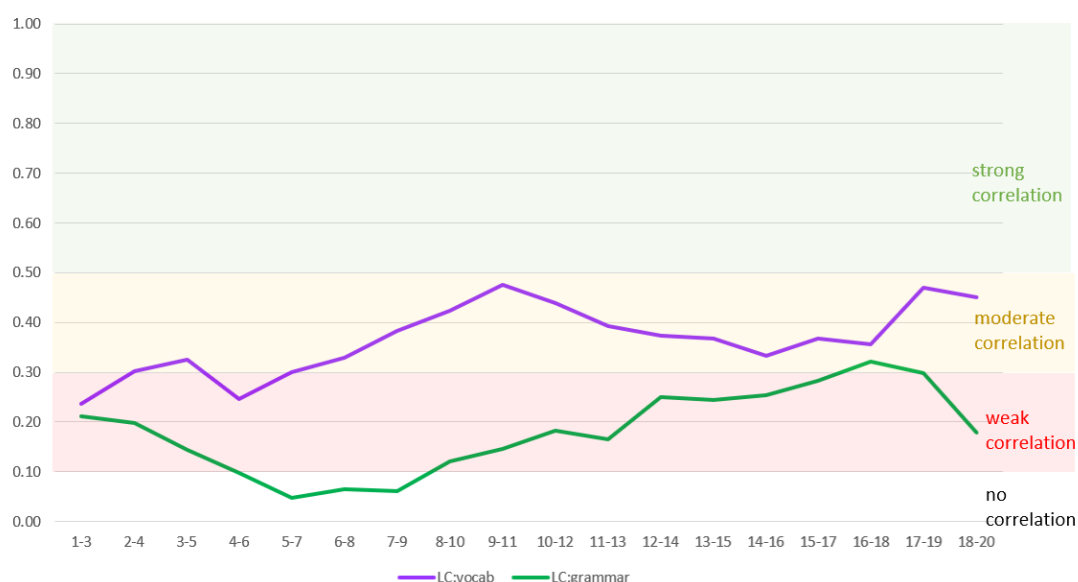


Figure 52: Moving correlations between listening and vocab / listening and grammar for the whole cohort

5.2.4. Linguistic subsystems: concluding remarks

The approaches demonstrated in the sections are common to CDST research which looks for variability to show different kinds of development (van Dijk et al., 2011). With regard to the individual subsystems, I have demonstrated that development is not linear. We have seen:

- A steady increase in listening proficiency throughout the data collection period, with the steepest increase initially. Six phases of learning were identified, with some suggestion that learning had stabilised around Time 12. However, after this the bandwidth again expanded, indicating a return of variability.
- A sharp initial increase in vocabulary size, levelling off around Time 7 and dropping again dramatically after Time 17. There was no evidence of a narrowing bandwidth, indicating that the mean score was consistently variable.
- Initially steep progress in grammatical knowledge, which continued to improve throughout the study, but less dramatically. Again, bandwidth fell around Time 12, suggesting that progress might have been maximised, but then widening, as with listening proficiency, demonstrating a return to flux.
- Vocabulary size and grammatical proficiency appear to grow together for much of the study period, but diverge from Time 16.
- Correlations between listening proficiency and vocabulary are consistently moderate throughout the study period, but correlations between listening proficiency and grammar are consistently weak.
- Sometimes these correlations appear to be in sync, at other times they compete.

5.3. Metacognitive subsystems, their development and relationships

The MALQ questions were asked at each time point, with the order randomised. The constructs evaluated have different numbers of questions; the six-point Likert scale was scored between 0 and 5, and the resulting number was transformed into a percentage and then put on a scale of 0 to 1, where the highest possible score of 1 represents ‘agree very strongly’ on all the possible points, and the lowest possible score of 0 represents

‘disagree very strongly’⁴⁶. The full MALQ questions can be found in appendix 15.1.1, but to summarise: Person Knowledge probes senses of anxiety and self-efficacy in listening; Directed Attention examines the levels of focus and concentration given during listening; Problem-Solving addresses the strategic use of context and co-text; and Mental Translation asks participants about the extent to which they translate as they listen. Note that at each time point, participants had been specifically asked to translate what they had understood, so that participants might have been primed towards higher numbers on this element, although as discussed in Chapter 4, they were also asked to transcribe, and as will be seen in Chapters 6 to 9, each of the case studies reported an inclination towards translation which seemed to be independent of the paused translation methodology.

5.3.1. Metacognitive scores

The full cohort mean scores with trendlines are shown in Figure 53.

The mean for the Person Knowledge construct sits consistently throughout the study in a band representing ‘agree slightly’, sometimes edging towards ‘agree quite strongly’. The Directed Attention mean sits in the same band until Time 17, where it falls slightly, indicating a slight inclination towards less focus, but still closest to ‘agree slightly’. The Mental Translation mean rises from an average closer to ‘agree slightly’ and approaching ‘agree quite strongly’ towards the end of the study, and Problem-Solving is closest to ‘agree quite strongly’ throughout the study period, with the exception of Times 1 to 3.

Moving min-max graphs are given for all four constructs in Figure 54 and are reported for consistency. However, in view of the six-point scale of the MALQ, in which 0.6

⁴⁶ 1 = agree very strongly; 0.8 agree quite strongly; 0.6 = agree slightly; 0.4 = disagree slightly; 0.2 = disagree quite strongly; 0 = disagree very strongly

represents ‘agree a bit’ and 0.8 represents ‘agree quite strongly’ it is clear that the change represented by the moving min-max graphs is very subtle.

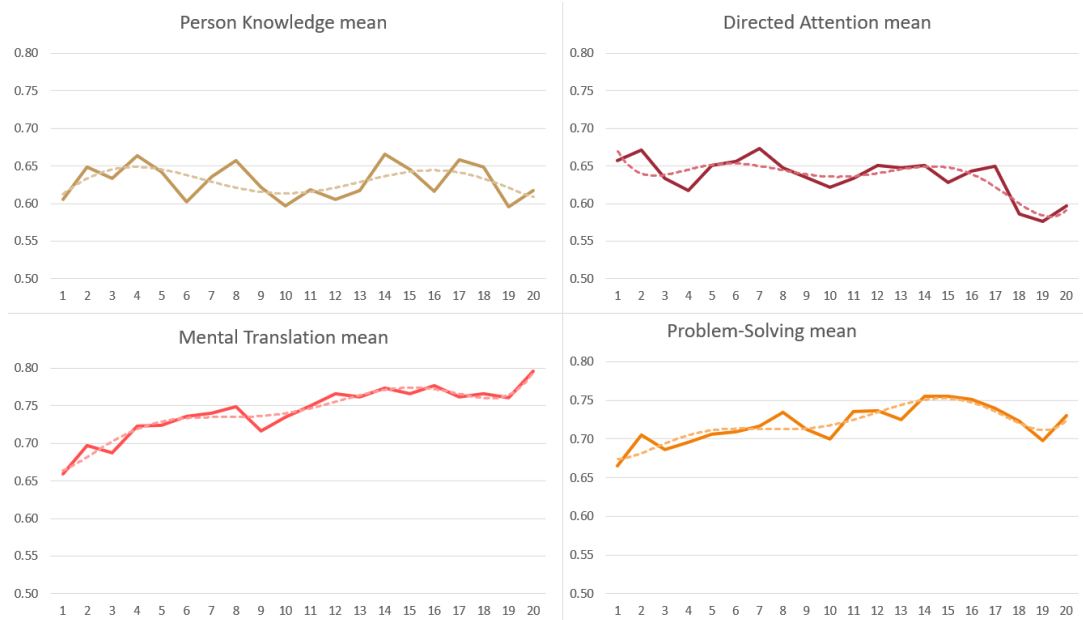


Figure 53: Means plus trendlines for the four MALQ elements

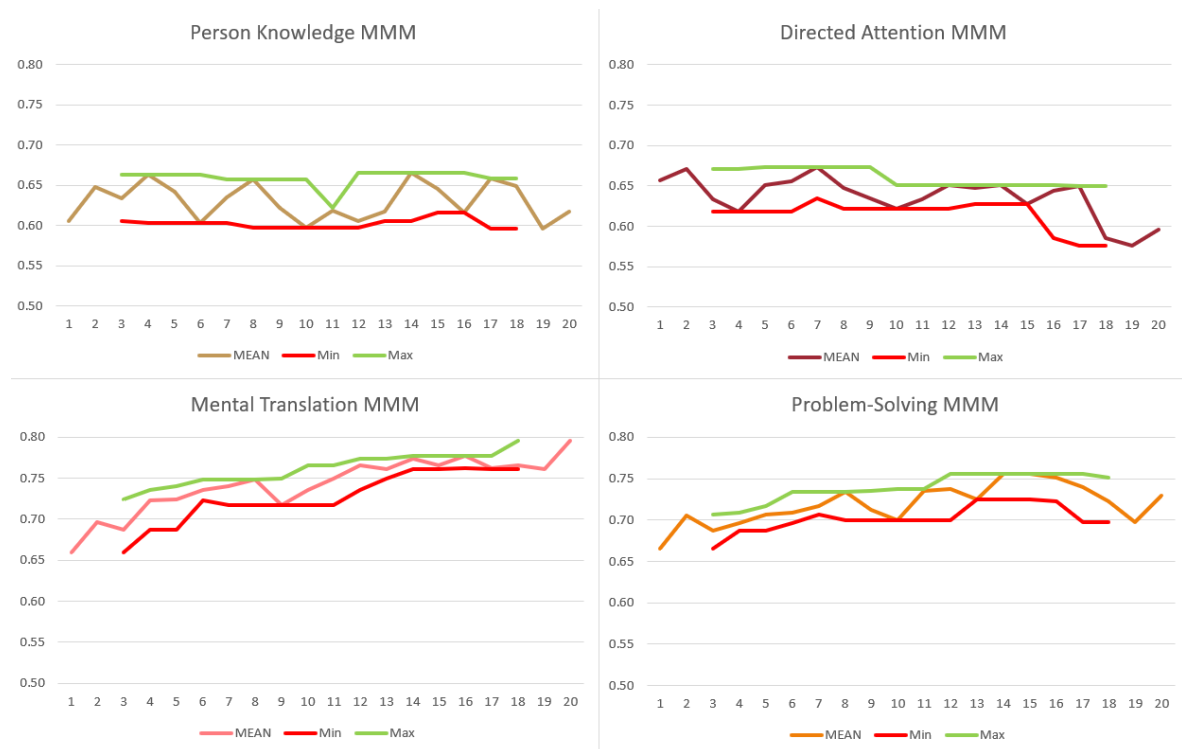


Figure 54: MALQ elements with moving min-max

5.3.2. Developing relationships between metacognitive subsystems

In a similar fashion to the linguistic subsystems, Figure 55 plots the four metacognitive subsystems on a single chart. To recap, the Problem-Solving statements related to participants' use of context and co-text to infer meaning from a partially understood passage, and Mental Translation referred to the extent to which participants translated as they listened. Statements relating to Person Knowledge relate to listening anxiety and participants' self-concept when listening, and Directed Attention addresses focus and concentration levels. The full Metacognitive Listening Questionnaire is set out in appendix 15.1.1.

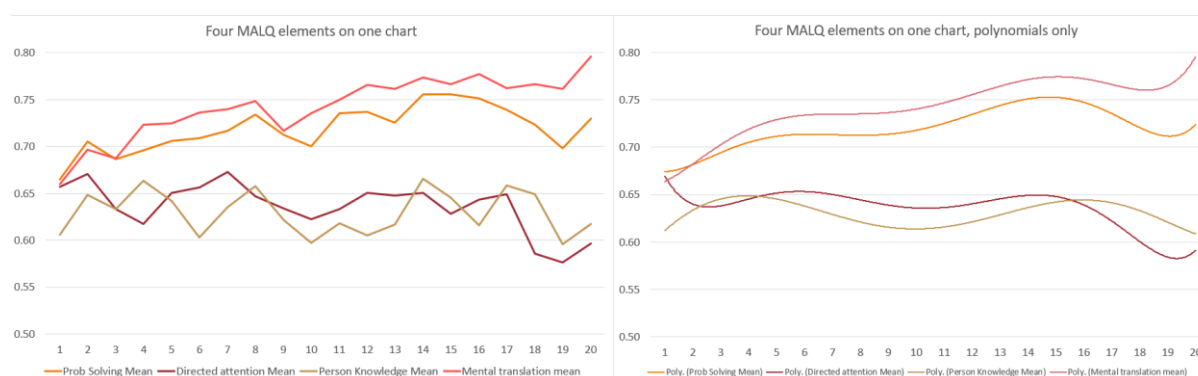


Figure 55: Basic MALQ relationships against each other (whole cohort)

Figure 55 shows a close relationship between the Mental Translation construct and that of Problem-Solving, and a relationship between Person Knowledge and Directed Attention that appears to begin competitively – in other words, initially listeners who are anxious and lack self-efficacy also report a lack of focus – but by Time 3, levels of anxiety appear to be broadly in sync with the attention paid. So we can see that throughout the data collection period, the more participants agreed with the Person Knowledge statements, the worse their listening scores were, although this was more pronounced earlier on in the data collection period and showed signs of improvement.

As far as Mental Translation is concerned, as discussed above, one might argue that the scores will be high given my conceptualisation of listening proficiency via paused translation. Nonetheless, it is notable that participants report that they are translating less by the end of the data collection period, suggesting that they might be beginning to build automaticity. Therefore it is reasonable to conclude these data are not entirely an artefact of the methodology.

The moving correlations between listening and the metacognitive subsystems of Problem-Solving, Directed Attention, Person Knowledge and Mental Translation are given in Table 44. Weak correlations are indicated in red and moderate in orange.

Table 44: Relationships between listening proficiency and the metacognitive subsystems

	L:PK	L:DA	L:MT	L:PS
1-3	-0.175	0.278	0.335	0.387
2-4	-0.231	0.223	0.426	0.326
3-5	-0.303	0.233	0.454	0.328
4-6	-0.329	0.172	0.392	0.245
5-7	-0.318	0.219	0.384	0.270
6-8	-0.274	0.284	0.399	0.294
7-9	-0.256	0.327	0.414	0.345
8-10	-0.177	0.296	0.436	0.352
9-11	-0.103	0.282	0.448	0.387
10-12	-0.097	0.279	0.437	0.448
11-13	-0.151	0.290	0.405	0.440
12-14	-0.187	0.283	0.435	0.419
13-15	-0.149	0.271	0.334	0.303
14-16	-0.200	0.302	0.315	0.301
15-17	-0.205	0.318	0.304	0.298
16-18	-0.217	0.360	0.371	0.343
17-19	-0.097	0.287	0.365	0.318
18-20	-0.038	0.283	0.304	0.355

We can see that in nearly all windows, there is a weak or moderate relationship, and these relationships are positive for Directed Attention, Mental Translation and Problem-Solving. In practical terms, this means that the higher the listening score, the more they agreed with the questions which related to each of these subsystems. By contrast, we also

see a mostly weak negative correlation between listening and Person Knowledge, which is associated with a lack of confidence and self-efficacy, and means that the more the cohort identified themselves to be nervous about listening, the lower their listening score. These moving correlations are also shown graphically in Figure 56, which highlights the largely unchanging nature of some of the constructs throughout the data collection period.

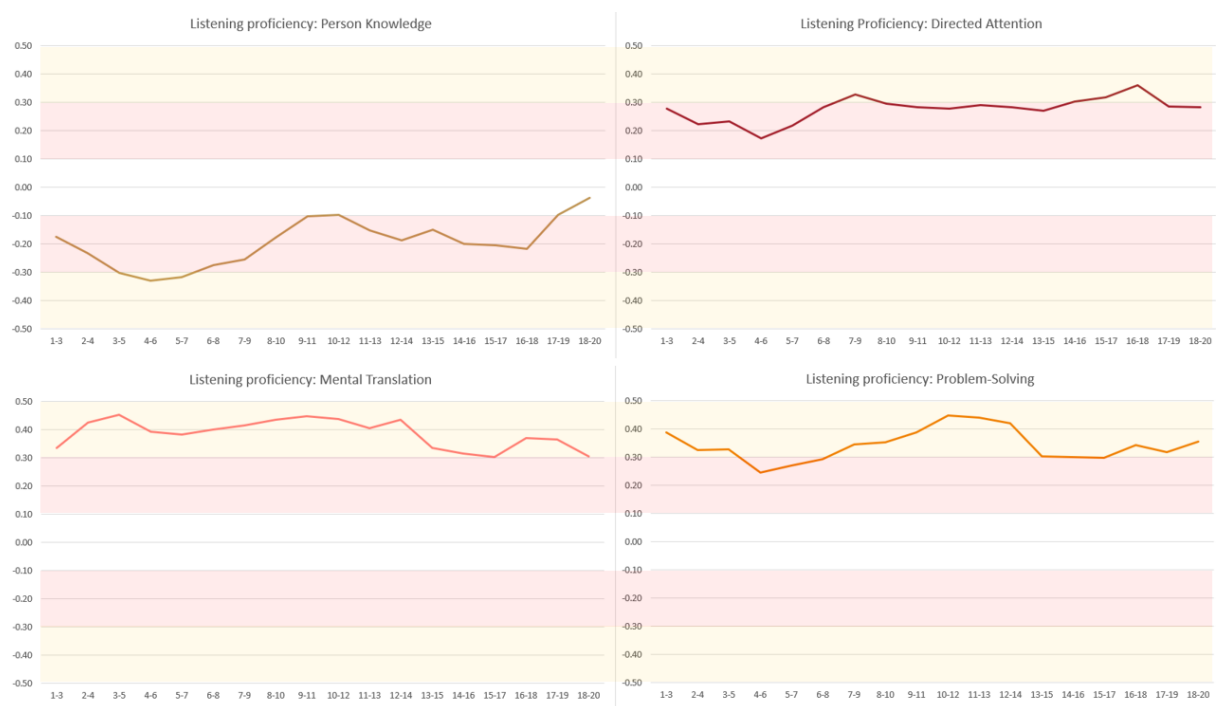


Figure 56: Relationships between listening and the metacognitive subsystems

We can see slight fluctuations in the correlation between listening proficiency and Problem-Solving, whereby the relationship strengthens marginally in the middle of study, but for the relationship between listening and Directed Attention, moving correlations stay remarkably level at around 0.3. There is more fluctuation in the relationship between listening and Person Knowledge, with the negative correlation between the two variables becoming stronger at the start of the data collection period and weakening after the Times 5-7 window, and ending within the ‘no correlation’ part of the graph. In other words, for

the whole cohort, in the final Times 18-20 window, the level of anxiety shown by the cohort had no relationship with their listening score. The relationship between listening and Mental Translation appears to be the most changeable, and, while it is the only one to stay within the ‘moderate’ correlation bracket throughout the study, shows a drop from the Time 12-14 window onwards.

5.3.3. Metacognitive subsystems: concluding remarks

- All four elements of the Metacognitive Listening Questionnaire had mean responses between ‘agree a bit’ and ‘agree quite strongly’, although the agreement was slightly stronger for Mental Translation and Problem-Solving than it was for Person Knowledge or Directed Attention. Only small fluctuations in bandwidth were visible.
- The elements of the MALQ can be split into two distinct pairs. Problem-Solving and Mental Translation evolve together, as do Directed Attention and Person Knowledge. The two pairs largely mirror each other.
- Moving correlations between listening proficiency and the elements of the MALQ remain fairly steady throughout the study period, with Mental Translation and listening proficiency showing the highest levels of correlation, and Person Knowledge and listening proficiency showing the weakest (negative) correlation.

5.4. Summary of findings 2

This chapter addressed the research question ‘how do the relationships between listening proficiency and the subsystems of language knowledge and metacognition develop over three academic years?’ To sum up, for the figures for the entire cohort we see a range of dynamic relationships which are working together within the full system of listening proficiency. These findings, taken together with the findings of the previous chapter

which examined qualitatively the types of errors made in the paused translation tests, shed light on the participants' experiences of listening from many angles. The findings will now be used to contextualise the experiences of the individual case studies in the subsequent findings chapters.

6. Findings 3: Helen

This chapter is concerned with the case study ‘Helen’. It explores her listening errors and what they might reveal about her listening processes and development, as well as how her subsystems and the relationships between them develop over time. By combining the first two research questions but answering them specifically to one individual, a picture can be drawn of the complex dynamics of Helen’s developmental trajectory, and examine the extent to which her experiences compare to those seen at group level.

6.1. Helen: Background information

Helen is a monolingual ‘higher prior attainer’ who lives with her parents and twin brother in the city. She studied Italian at primary school and reported that no-one else in the family spoke any French. She reported minimal out-of-class exposure to French with the exception of some short episodes on the Duolingo app and a summer holiday to France between Time 6 and Time 7. Particularly in her first year of secondary school, Helen was a nervous girl with a quiet voice, who often talked in our interviews about her anxiety and desire to excel. She took her role as a case study seriously and frequently discussed classroom episodes of which she had made a mental note because she thought they would be of interest to me, and showed insight and reflection in our conversations. She took care over her answers to the surveys in order to help me, saying at Time 19,

‘I feel like if I wasn’t being interviewed, I would have done it [the listening survey] quicker. I wanted to go slow so I could remember it.’

Helen was highly strategic and began to interpret the questions in the MALQ as ‘best practice’, expressly stating at Time 8 that she was using them for ideas of how she could

improve her French listening. She consistently scored in the top quartile for listening proficiency, and continued with French after the study ended.

Throughout the study Helen claimed to be highly motivated by success, although in her second year of study she began to voice more dissatisfaction with school: following a change in class grouping she was no longer with her friends and on several occasions reported the boys in the class as ‘annoying’. In addition to this she was seated beside a friend with special educational needs and saw it was her duty to help her, which, she felt, detracted from her own learning. The upshot of this was that by Time 10 (the fourth time point in her second year of learning), she put French at seventh favourite school subject out of ten. Things started to look up at Time 11, when she celebrated both her thirteenth birthday and her parents’ wedding.

The periods of lockdown followed very shortly after this (Times 12 and 13 in lockdown; Times 14 to 16 lessons in school, but with data collection for the present study done at home, T17 in lockdown again, T18 to 20 lessons and data collection at school): Helen felt that her workload had increased and at Time 17 reported that online French lessons consisted of ‘*looking at slides and answering questions*’. However, on reflection at the end of the study she felt that this was a turning point: the lack of interactivity in online lessons forced her to explore other ways to learn:

‘I thought, “this might impact my learning if I just don’t do anything, and I actually want to learn French.”’

She said that in Year 7 (Times 1 to 6), she saw French as ‘*a lesson*’, but by Time 19, it was a language tied to countries that she wanted to visit, and underlying cultures to explore.

6.2. Helen's listening proficiency and its variability

6.2.1. Variability in listening proficiency scores

Helen's correlation co-efficient between listening proficiency and time point was 0.93 (the highest of all the case studies), demonstrating a very clear link between time and listening score. Her score at Time 1 was three points above the cohort mean, but her progress was steadily in advance of the mean, resulting in a Time 20 score that was 23 points above the mean. Helen's final listening proficiency was equivalent to a CEFR level B1, whereby she demonstrated understanding of straightforward factual information about common topics and cope with the main points that she heard in short narratives.

Table 45 and Figure 57 give Helen's raw listening scores drawn from the Rasch model. The changes in listening proficiency plotted on the moving min-max graph (Figure 58) reinforce the impression of steady progress of fairly small increments (relative to the other case studies) and small windows of variability. When Helen was progressing in her listening, she did so confidently, and without the large fluctuations which were seen in the case studies of Theo and Ryan.

Table 45: Helen's listening proficiency compared with the mean

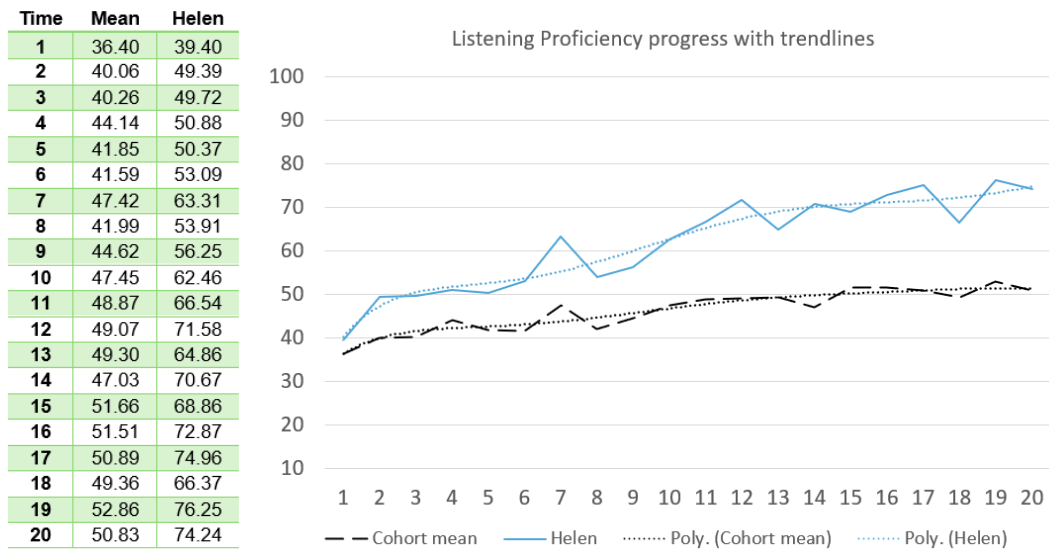


Figure 57: Helen's listening score compared with the mean, including trendlines

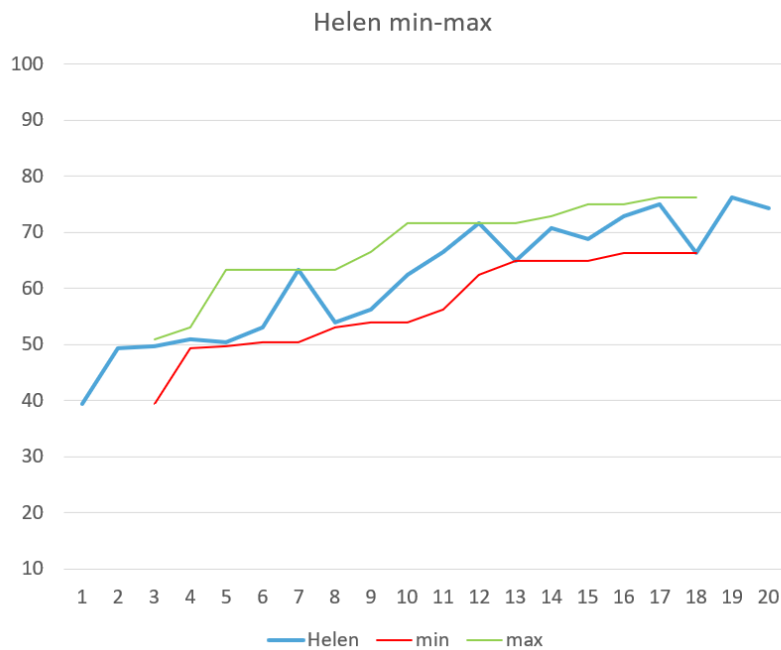


Figure 58: Helen's moving min-max listening score

The change point analysis software (Taylor, 2000) reveals two key change points during her developmental trajectory, at Time 7 and Time 11 (Figure 59). Both changes occurred with 95% confidence and the level associated with the change indicates its importance:

the level 1 change was detected on the first pass through the data, and the level 3 change was detected on the third pass. The blue background represents a region expected to contain all the values based on the current model, and the red lines are control limits representing the maximum range over which the values are expected to vary. For a reminder of the procedures involved in the CPA, refer to methodology section 3.6.3.

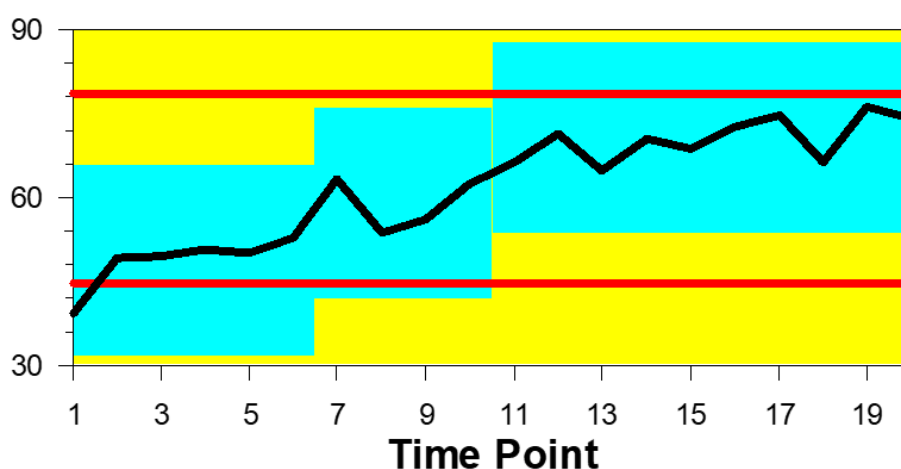


Figure 59: Helen's listening proficiency changes according to change point analysis

6.3. Helen's linguistic subsystems, their development and relationships

6.3.1. AuralLex scores

Helen's vocabulary scored as measured by the AuralLex test, compared with the mean, are given in Figure 60. Her scores remained lower than the mean throughout the first year of the study. At this point she began to make rapid progress which peaks at an estimated vocabulary size of 850 words at Time 19, although with fluctuations visible at Times 15, 18 and 20.

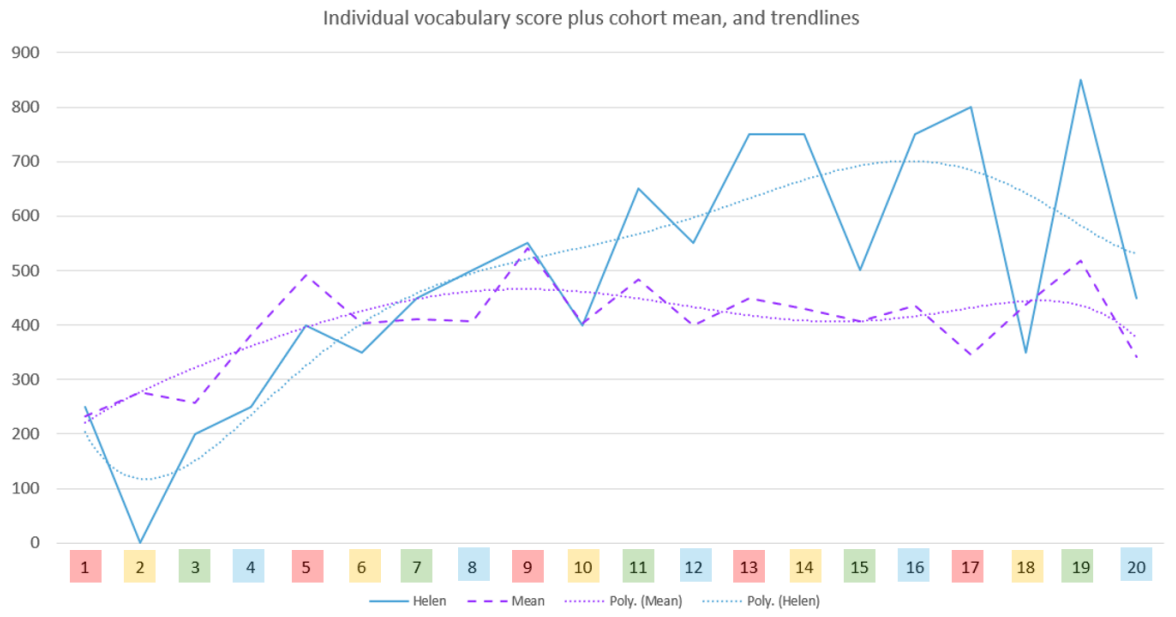


Figure 60: Helen's AuralLex scores (compared with mean)

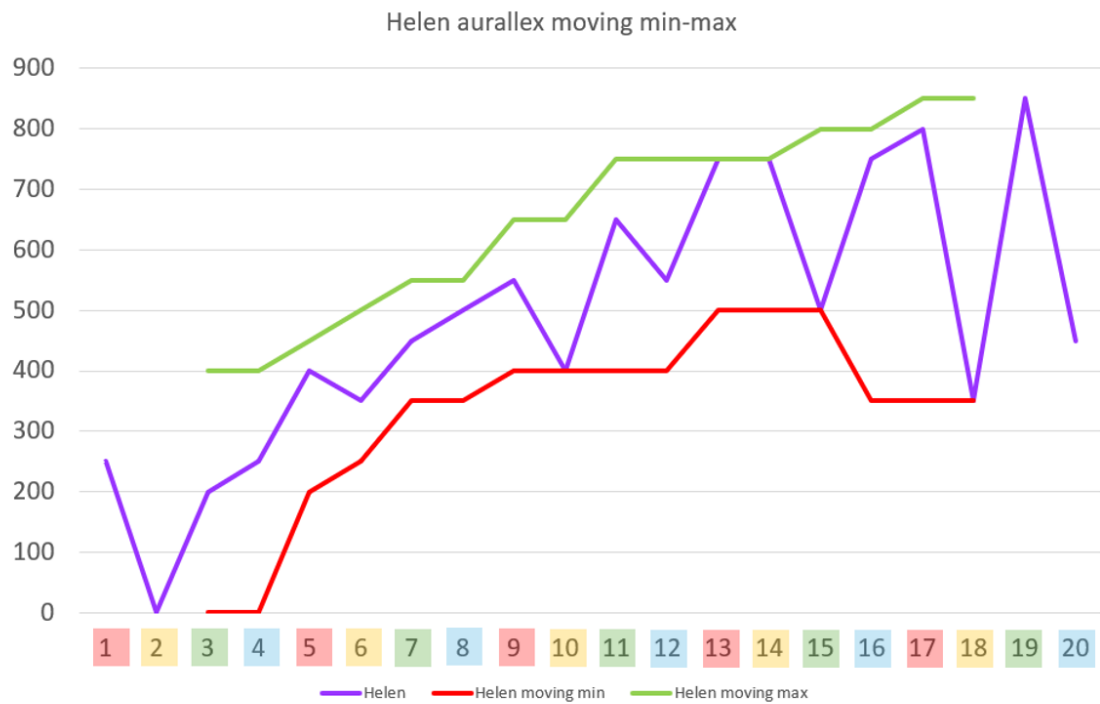


Figure 61: AuralLex moving min-max for Helen

When these AuralLex scores are viewed through the lens of the moving min-max graph (Figure 61), the dramatic decrease from Time 15 is clear. It might be accounted for by the nature of the instrument, with high penalties for incorrect guessing: as Helen’s confidence grew (see Figure 69), she began to incorrectly identify French non-words, which was rarely a feature of her AuralLex responses in earlier tests. These fluctuations in scores between time points, (or wider bandwidths), when viewed from a CDST perspective, can be seen as indicative of a path of progress, whereby a period of fluctuation frequently occurs before a jump in progress; had the data collection continued for a few more time points a new, higher level might have emerged.

6.3.2. AuraGram scores

Helen’s grammatical development as measured by the AuraGram is shown in Figure 62. We can see that despite her above-average progress in listening proficiency, her performance was below the average until Time 7. She consistently exceeded the average only from Time 14 and by the end of the study her scores had again reverted to the mean.

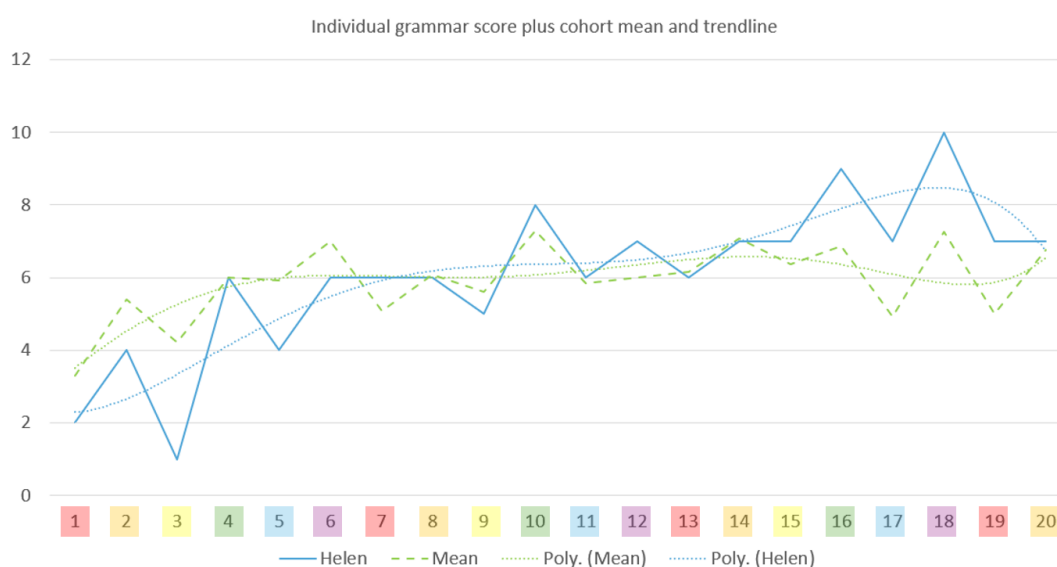


Figure 62: Helen's AuraGram scores with reference to the mean, plus trendlines

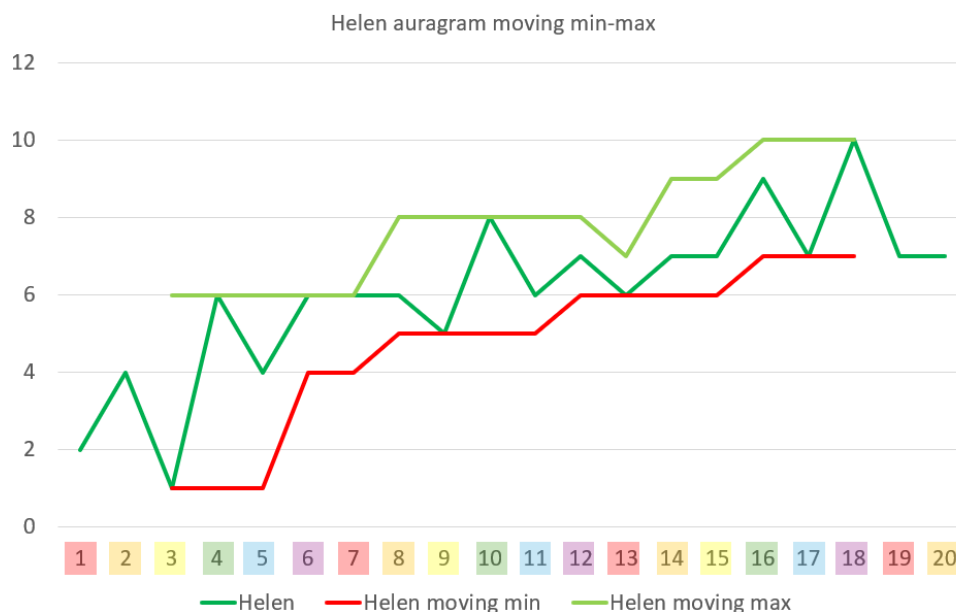


Figure 63: Helen's AuraGram moving min-max

Helen's moving min-max scores are shown in Figure 63, highlighting fluctuations in bandwidth and illustrating erratic performance. The narrowing of the bandwidth at Time 13 is suggestive of her approaching the possibility of some stability in grammatical knowledge, but this was temporary, and she then began to progress again, with her highest score achieved at Time 18. In CDST terms, this narrowing or semblance of a plateau is known as an 'attractor state' – the system is attracted to this kind of state and is likely to remain in it for some time. A particularly large push might be needed for the system to begin to move on again.

Statements about her knowledge of French grammar made at interview contradict the quantitative data. Relative to her vocabulary knowledge, she said at Time 15 that '*I still wouldn't know if the grammar was right*', and when asked to reflect on the entire study in the closing Time 20 interview, she stated that French grammar was '*still quite*

complicated’ and referred to her learning of grammar as one of the low points of the last three years of French study.

6.3.3. Developing relationship between linguistic subsystems

In order to address the relative roles of lexical and grammatical knowledge in Helen’s developmental trajectory of listening, Figure 64 places Helen’s vocabulary, grammar and listening scores together on the same graph, normalised on a scale of 0 to 1 for ease of comparison. As stated in section 5.2.1 of Chapter 5, for context, in vocabulary terms, the 0.5 line represents a vocabulary size of 600 words and 0.25 represents a vocabulary size of 300 words (as the maximum observed score across all time points was 1250 words). In grammar terms, the maximum observed score was 15/20, so the 0.45 line represents a score of 7/20.

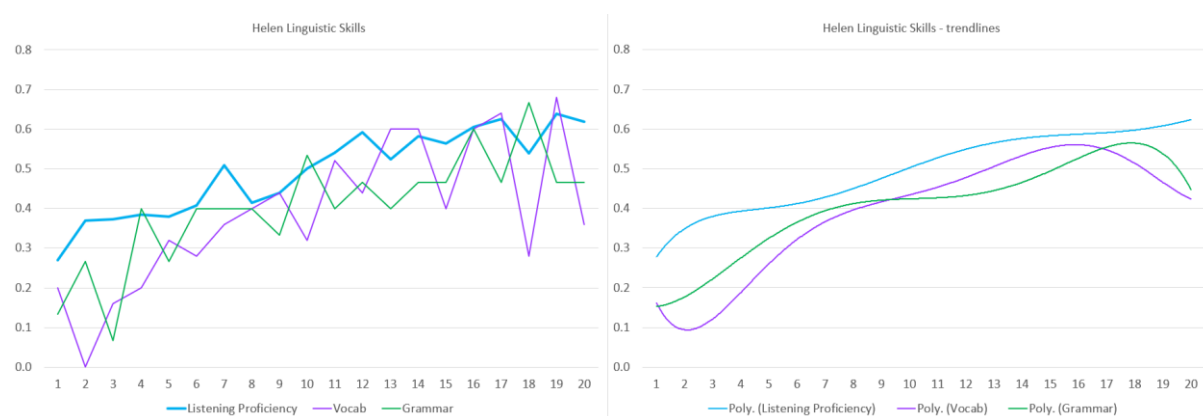


Figure 64: Vocabulary and grammar charted together

Moving correlations between listening proficiency and both vocabulary and grammar are given in Figure 65. They indicate a highly variable relationship which, in the main, could nonetheless be categorised as ‘competitive’: when the correlation between grammar and listening was strongly positive, the relationship between vocabulary and listening was strongly negative, which might suggest that Helen only had the capacity to use one or other of these subsystems to help her with her listening at any one time.

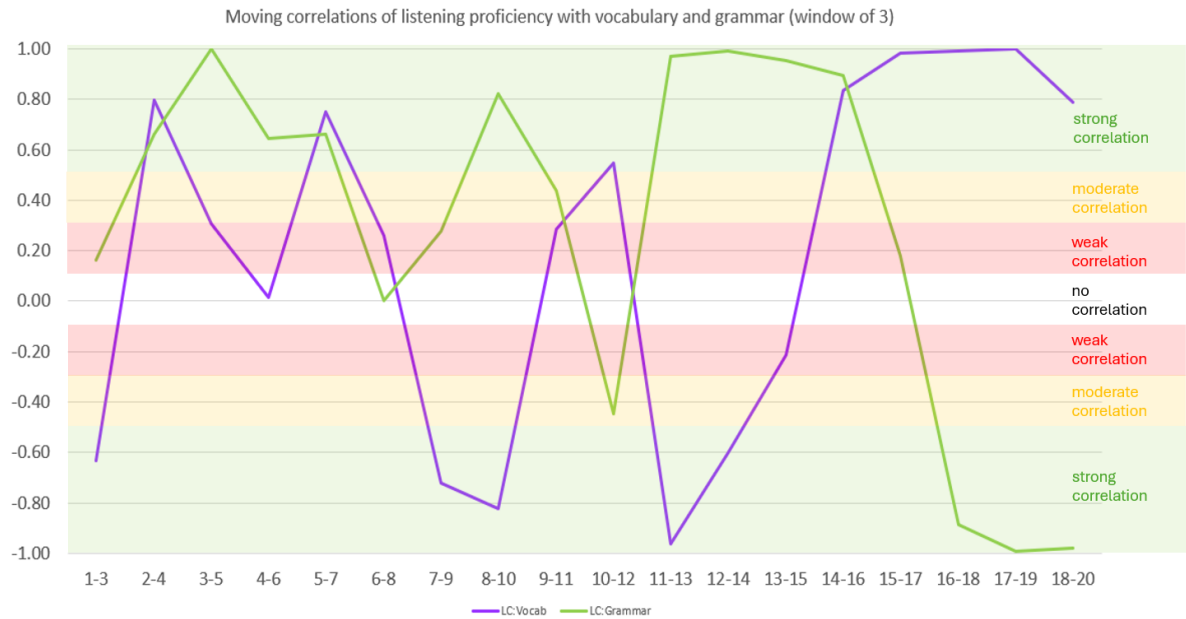


Figure 65: Helen's moving correlations of listening proficiency with vocabulary and grammar

6.4. Helen's errors and listening processes

Above I examined Helen's developmental trajectory from a descriptively quantitative viewpoint, comparing her experience with those reported in Findings 2 for the whole cohort. I will now explore Helen's personal listening trajectory, and present them in relation to the findings discussed above. Findings 1 roughly divided errors into those which were phonologically driven, and those which were textually driven, while acknowledging that the complex and dynamic nature of the moment-to-moment experience of listening did not always lend itself naturally to this dichotomy. To compare these whole-cohort findings with Helen's individual experience, I draw both from her paused translation data, but also from her interview data, which included stimulated recall of one paused translation exercise at each time point.

6.4.1. Perception of speed

Helen's perception of the speed of input depended on her level of understanding: where she struggled to understand the French, her interpretation of this was that the delivery was accelerated, and although by Time 11 she was often demonstrating good understanding, she reported that there were still times when *'it goes too fast'*. In fact, the input had been carefully calibrated (see methodology) to be consistent in terms of speed (measured in words per minute). Her perception of speed also seemed to relate to the amount of listening she had been doing: on return to school after the end of the first covid lockdown, for example, at Time 15, she said that listening *'felt faster'*. With time she came to acknowledge the delay in processing French, by stating at Time 18,

'You know how in English if someone says something really rushed and then you know what they said a second later – that's sort of what I do with French.'

By Time 20 Helen had recognised the link between reduced understanding and a perception of speed, indicating growing insight into her own listening experience. This insight will be examined further in section 6.6 on the development of her metacognitive processes.

6.4.2. Favouring consonants over vowels

In Chapter 4 section 4.3, I discussed at length the issues whereby English words were erroneously detected within the French stream of speech, yet this was not a listening error that Helen demonstrated. Rather she was inclined to erroneously detect French through an emphasis on consonants and a disregard for vowels. In order to attach some appropriate meaning she sometimes overlooked word boundaries. An example of this was cited in Findings 1 where *mais c'est beaucoup* /mɛ sɛ boku/ ('but it's a lot') was

translated as ‘thank you very much’⁴⁷ (*merci beaucoup* /mɛʁsi boku/): at Time 9 Helen was one of the participants to make this error. Her translation ‘thanks a lot you three’ was rooted not only in a conflation of /mɛ sɛ/ and /mɛʁsi/ but also exacerbated by focussing on the /tʁ/ of ‘trop’ to give it a meaning of ‘three’ (‘trois’ /tʁɑ/) rather than ‘too much’. However, while the ‘merci beaucoup’ problem persisted for many participants at Times 13 and 18 (6/37 at T9, 4/31 at T13 and 10/37 at T18), Helen made this error only at Time 9.

Although it is tempting, then, to explain this improvement in relation to the change point identified at Time 11, Helen made a range of other errors rooted in her favouring of consonants over vowels throughout the data collection period. These include ‘I love’ (*j’adore*) for ‘j’ attends’⁴⁸, ‘favourite’ (*préféré*) for ‘I prefer’ (*je préfère*)⁴⁹, ‘you’ (‘tu’) for ‘tout’ (all)⁵⁰, ‘my mum’ (*ma mère*) for ‘mais moi’⁵² and ‘in Italy’ for ‘on est allés’⁵³. It was also apparent that Helen overlooked vowel sounds during our Time 8 interview where she heard the word ‘vélo’ /velo/ and instantly repeated it back as /velu/.

In section 4.3.3 of Chapter 4 I discussed at length the confusion experienced by participants between the words *camping* / kɑ̃piŋ / (‘camping’ or ‘campsite’), *campagne* / kɑ̃paŋ / (‘countryside’) and *copains* / kɔpɛ̃ / (‘friends’). Helen experienced this confusion on multiple occasions and explained it clearly at Time 11:

⁴⁷ track 8 / test D25 / T9, T13, T18

⁴⁸ T3

⁴⁹ T3

⁵⁰ T10

⁵¹ T20 tous mes vêtements

⁵² T15

⁵³ T18

'and then I heard camping, but at the start I thought was, like, a friend, like 'copine', because my brain just instantly said 'friend' because I hear that more than 'camping'.

In fact, her allusion to *copine* /kopin/ rather than default masculine form *copain* also indicates her dependence on consonants to give meaning to words, and her overlooking of vowels.

At Time 14 she showed uncertainty in response to *on a une maison à la campagne* ('we have a house in the countryside') acknowledging a confusion as to why a 'house' (*maison*) should be referred to within the context of camping, but explaining it away. And at T18, in response to *on ira à la campagne. J'aime bien la campagne* (we'll go to the countryside. I like the countryside'), she said, '*I didn't know whether they were talking about camping or friends*'. At no point in the data collection period did Helen show any knowledge of the word 'campagne' meaning countryside.

What is striking, though, in all these examples of Helen's phonological difficulties with interpreting French, is that unlike the cohort more widely, her reference is always to other French words rather than finding non-existent English words within the French stream of speech.

6.4.3. Sound and sense work together

In fact, while the cohort at large tended to favour the sound of the input even at the expense of the sense of any translation, this was not the case for Helen, for whom sound and sense interacted in order to build meaning, even if in some cases this meaning was

incorrect. For example, in response to the input *moi, j'habite dans une maison individuelle* ('I live in a detached house')⁵⁴ she said in stimulated recall:

'It said 'moi, j'habite' and that's, 'me, I live in', and then it's, 'my house', and then the something that sounded like "DVD" or something. I didn't know what that was, so I just left it.'

In section 5.3.2 I discussed the rare but noteworthy phenomena of metathesis (misordering of syllables). Helen provided an example of this, at Time 14, but again ensured that her translation made sense within the context that she had understood. When translating *J'adore aller chez ma cousine* ('I love going to my cousin's home'), she said 'I love the food / cuisine.' She told me that her translation was lead by understanding *cousine* as 'cuisine' (another example of favouring consonants over vowels).

6.4.4. Content and function words

The balance between wrong and missing words with respect to content and function words was discussed for the sample as a whole in Chapter 4 section 4.5.1: there, I reported that the proportion of missing words was fairly evenly distributed for function and content words; however, content words were much more likely to be wrong, but function words much more likely to be omitted. This was the case with Helen just as it was with the full cohort: this is consistent with the caution she showed in her answers in the dataset. I had made two hypotheses about this in Findings 1: firstly, a deliberate overlooking of some function words in order to build a meaning and secondly, skipping function words at the expense of content words.

⁵⁴ track 2 of D47 (heard at T11 and T17)

In terms of the first of these hypotheses, I gave the example of *j'habite à la campagne* being translated by many as 'I live with my friends', the translation forcing the inclusion of the word 'with' which is absent from the input. Helen's experience was not consistent with this, however. For example, when struggling with *je pars avec mes amis* ('I go with my friends'⁵⁵), the very presence of the word *avec* ('with') led her to translate the sentence as 'I speak with my friends'. In explaining her response she stated:

'I couldn't tell whether he said pars or parents, but I tried to use logic. So I assumed with friends, he's speaking with his friends. Then I started doubting myself, I thought did he say parle or did he say something I haven't heard of before?'

In other cases, Helen's struggles with prepositions appeared to epitomise the cohort's experience of skipping function words at the expense of content words. She appeared sometimes to be lulled into a false sense of security when she perceived a sentence to be particularly easy. For example *je joue souvent au foot* ('I often play football') was translated as 'I like to play football', which she stated as 'easy', even saying '*when I replayed it in my head I was like, "oh yeah"*'. At other times she recognises the inherent difficulty with prepositions: in response to *Bien sûr, j'ai un portable* ('of course I have a phone')⁵⁶, she states:

I just heard bien, and then, sur is one of the shorter words that just... I don't know. Sometimes with shorter words they just sort of go under the radar. So I think – I'm not really sure what the answer means because I thought sur was

⁵⁵ T14

⁵⁶ T19

'on'... then sometimes I skip out the little words. I acknowledge that they're there, but I don't always focus on them then I'll miss the bigger words.

6.4.5. Morphological features

In Findings 1 I discussed participants' failure to recognise morphological features and hypothesised that this was unsurprising given that function words were themselves hard to recognise, and morphological features are effectively 'function morphemes', even smaller and less discernible than whole words. Helen's profile was not dissimilar to that of the whole cohort on average. For example, Figure 41 in section 4.5.2 (copied here in Figure 66) illustrates the whole cohort's struggles with plurality over time, indicating that although there was improvement with the recognition of the singular 'un film', there was not a commensurate recognition of the plurals.

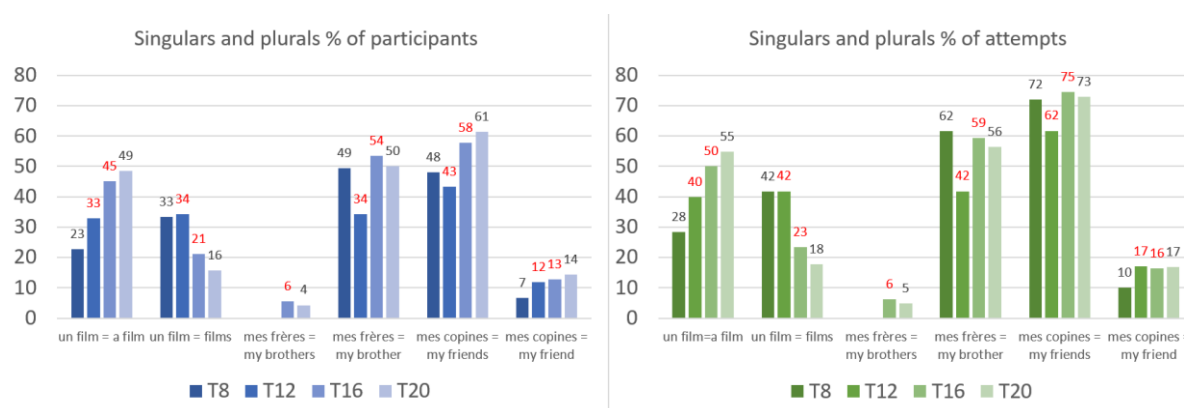


Figure 66: Grammatical errors of plurality in test D8

Helen's own data demonstrated an unstable grasp of morphological features indicating plurality. At Time 8 she was unable to translate the phrase *je regarde un film d'épouvante avec mes frères* ('I watch a horror film with my brothers') at all. At Time 12 she spoke of 'films with my brother' and then in track 7 referred to *mes copines* ('my friends') as 'my friend' – in other words, making all the possible errors she could. By Time 16 she correctly translated *un film* as 'a film', rather than 'films', but did not identify either of

the subsequent plurals, and at Time 20 she reverted to ‘films’, correctly identified *mes frères* as ‘my brothers’ but incorrectly identified *mes copines* as ‘my friend’ in the singular. During stimulated recall in a similar case at Time 19 I asked her about this, and she attributed her choice of translation both to transfer of training, as well as overlooking the minutiae of the details:

‘I think in class we did examples where it’s like ‘my friend’ rather than ‘my friends’ So it could have been my friends, I just... And also because mes is a short word, so sometimes you kind of scan it in French.’

In fact, there was a wealth of evidence of Helen’s ignoring of morphemes from across the dataset, and no suggestion that this was an element of her listening proficiency that developed with the change points: in test D25, after a successful translation of *j’aime regarder* as ‘I like to watch’ at Time 9, she reverted to ‘I’m going to watch’ at Times 13 and 18. At Time 11 she translated *nous* (‘we’) as ‘I’, and a similar erroneous default to the first person was seen at Time 15. In Findings 1 I attributed the overuse of the first person to transfer of training: here it coincided with the lack of salience of some elements of the input.

6.5. Summary of findings for the linguistic systems

We have seen:

- A steady increase in listening proficiency throughout the data collection period, (T1 listening score 39; T20 listening score 75) with change points statistically identified at Times 7 and 11, underpinned by Helen’s perception that input was slower after Time 11: a perception that seems to be a proxy for better understanding.

- A growth in vocabulary size which peaked at Time 17, but with very large fluctuations from Time 15.
- A relatively large vocabulary size and no inclination to erroneously detect English words within the French input, unlike the cohort at large.
- Variations in grammatical score throughout the study with a suggestion of growth overall, but without the variations seen in vocabulary.
- Moving correlations which suggest a competitive relationship between vocabulary and grammar.
- Helen's perceptual experience of listening conflated speed of delivery with lack of understanding.
- Unlike the cohort as a whole, sound for Helen did not 'trump sense', but instead she exploited context to build meaning.
- Qualitative data reflected her grammatical struggles with function words and function morphemes.

6.6. Helen’s metacognitive subsystems, their development and relationships

6.6.1. Metacognitive scores

Helen’s scores for the metacognitive subsystems are given in Figure 67, relative to the mean, with moving min-max scores in Figure 68. Recall that the statements were rated on a six-point Likert scale ranging from ‘strongly agree’ to ‘strongly disagree’, where ‘strongly agree’ would equate to a score of 1.0. The statements on the MALQ are in appendix 15.1.1.



Figure 67: Helen’s metacognitive scores relative to the mean

The moving min-max graphs in Figure 68 show decreases in Person Knowledge and Mental Translation scores, (indicative of reducing anxiety and less translation) which align with the both change points reported in section 6.2.1. The Directed Attention and

Problem-Solving moving min-max graphs also suggest changes at Time 7, aligning with the first of the change points.

Where the findings do not suggest changes in Helen’s linguistic subsystems at Time 7 and Time 11, the data drawn from the moving min-max graph suggest that in Helen’s case, her metacognitive subsystems align more closely with her ultimate listening proficiency. This will be explored in more detail through both the moving correlations and the many insights that she gave as to her listening metacognition through the interviews that took place at each time point (for interview protocol see appendix 15.9 and for the transcript of a typical interview, see appendix 15.9.1).

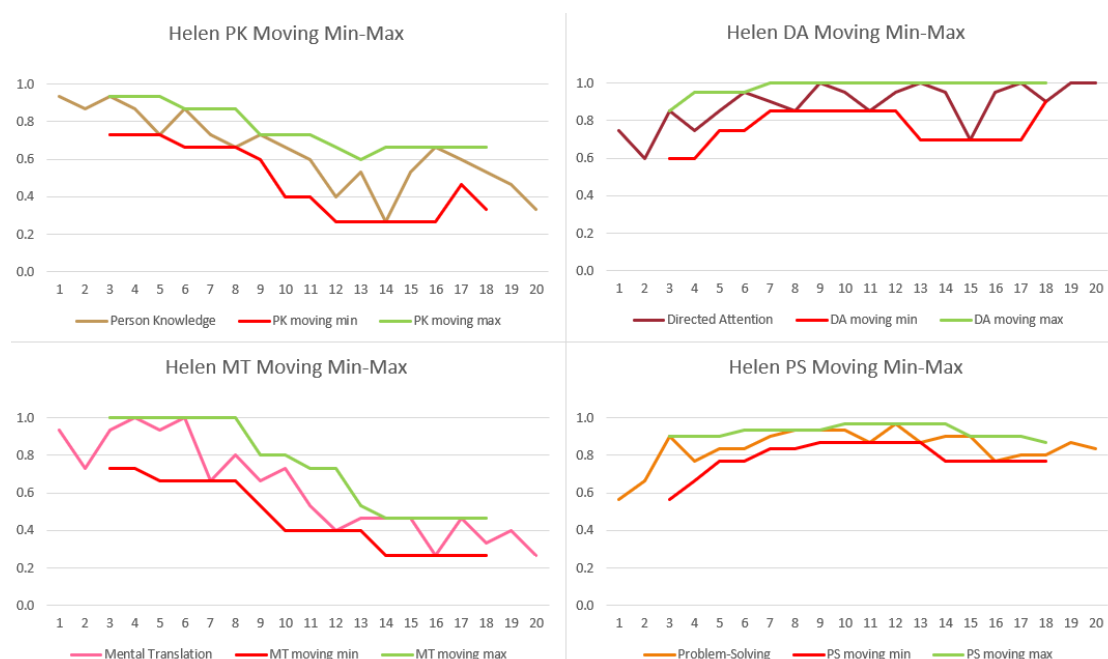


Figure 68: Moving min-max for Helen’s metacognitive scores

6.6.2. Developing relationships between metacognitive scores

Recall that in section 5.3.2 of Chapter 5, I showed that on average for the whole sample, trajectories of Mental Translation (pink) and Problem-Solving (orange) followed each other, and that the same was true of Directed Attention (burgundy) and Person Knowledge (beige). In other words, students who were less anxious were more likely to

pay better attention to the input, and Mental Translation worked in unison with use of context and co-text to build understanding.

Figure 69 illustrates the development of Helen’s metacognition, and shows a very different pattern. For her, a reduction in Mental Translation aligns with a reduction in anxiety (measured through Person Knowledge), while her Directed Attention trajectory and that of Problem-Solving also largely follow each other. For ease of interpretation I have set out the moving correlations in two pairs in Figure 70 to reflect the relationships illustrated in Figure 69.

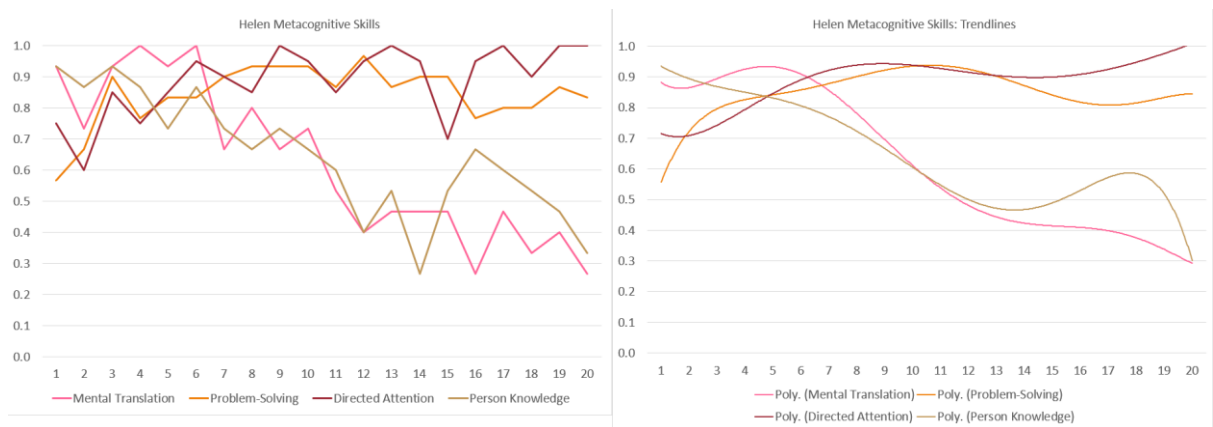


Figure 69: Development of Helen’s metacognitive skills

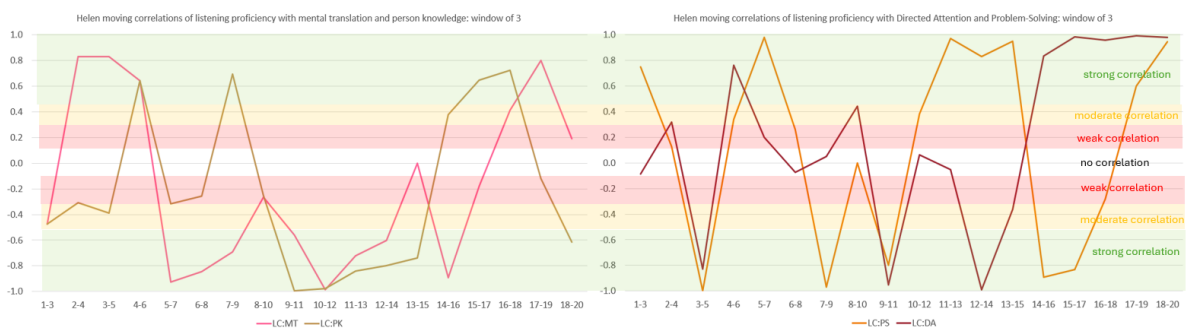


Figure 70: Moving correlations of metacognitive subsystems viewed together

By superimposing Helen's moving correlations of Mental Translation with listening proficiency on that of Person Knowledge with listening proficiency, I make an inference about the relationships between the subsystems as well as each subsystem's relationship with Listening Proficiency. We see a relationship where initially Mental Translation was a precursor to Person Knowledge; then the two are briefly in competition aligning for an extended period of time. Finally the correlation between listening proficiency and Person Knowledge reverted to strongly positive, with Mental Translation following a very similar pattern one window later. The moving correlations between listening proficiency and Problem-Solving and Directed Attention began supportively, before moving briefly to compete with one another and then reverting to a supportive relationship. However from the Time 10-12 window onwards, the relationships were in competition: when Directed Attention scores correlated positively with those of listening proficiency, Problem-Solving correlates negatively, and vice versa. In other words, Helen's good listening scores were related *either* to paying good attention to what she heard, *or* to a careful use of Problem-Solving techniques, but never both.

6.7. Helen's strategic journey

In this section I will discuss Helen's insights into her metacognition based on what she told me in interviews at each time point. I will separate this into the three phases identified by the change point analysis software.

6.7.1. Phase 1: An anxious mental translator determined to succeed

In the first phase identified by the change point analysis software, Figure 70 identifies a strong moving correlation between listening proficiency and Mental Translation. This is a pattern which was followed two time points later by a similar pattern in the correlation between listening proficiency and Person Knowledge.

Helen began her listening trajectory mentally translating everything that she heard: she told me this at interview, aligning with what she had reported in the MALQ questionnaire. This is not simply a result of the paused translation methodology of the study: she volunteered that she found dictation (i.e. paused transcription) activities in class difficult because she would listen to the French, would not be able to resist translating it into English in her head (*'It's just something I do'*), then would have to translate it back into French to write her answers for the task. This translation process would slow her down: she acknowledged that during the test battery, she would miss subsequent words because she was *'too busy translating'*, and even at Time 6 she referred to translating as *'a natural instinct'*.

With regard to Person Knowledge, Helen expressed her determination to succeed from the outset, but her assertions at Time 2 and Time 3 reinforced the moving correlation finding that her Person Knowledge score had a strong negative correlation with her listening proficiency – for example: *'I want to be good at it'* (Time 2) *'I was in a panic in the test'* (Time 3) and *'I feel determined to get it right but also disappointed I don't know what the words mean.'* (Time 3). Yet by Time 5 she was reporting that her listening anxiety is subsiding. This is at the high point of the positive moving correlation between her Person Knowledge score and her listening proficiency, and she equates proficiency with calm:

'I thought I did better than before. Because I felt more calm even when I didn't know some of the words. I had more confidence I could do it.'

An interaction between Problem-Solving and Directed Attention is also clear in this phase not only from the moving correlation, but also from Helen's interviews, and can be summarised by the idea that Helen's effort is sometimes more directed towards thinking

about the work than actually doing the work. At Time 1 she has identified a gap-filling strategy:

‘I hear a couple of words and then I think, “how would I make that into a sentence that makes sense?”

– an approach she refers to again at Times 2, 3 and 4 but it seems that during the early period of her developmental trajectory, this strategy had the potential to obscure the actual task:

‘I want to do well, then I get scared. I can’t separate the words, or I just panic and it all goes into a blur. And when I panic, I forget some of the meanings of the words, like when you’re in the middle of dropping something but you can’t catch it because it’s... well, it’s bad because I lose my focus and then I’m just thinking ‘oh my gosh no, I’ve done wrong’ and then I’m not actually listening to the French.’

6.7.2. Phase 2: Helen as a ‘self-subtitler’

At Time 6 Helen told me,

‘Usually when I’m listening – especially in French – I can see the words in my head.’

This act of self-subtitling became a theme in our interviews. She reported that visualising words seemed to override her translation instinct. On two occasions she gave lucid accounts of the experience of this ‘visual listening’. At Time 7, we had been talking about the translation of *j’aime habiter ici parce qu’il neige beaucoup*. (‘I like living here because it snows a lot’), and she translated the sentence with the exception of *neige*

(‘snows’). We had moved onto talking about something else, when she suddenly came back to the word and told me its meaning.

‘It’s not so much like subtitles, it’s just like, the words I’m more confused about.’

So talk through what happened with neige just then.⁵⁷

‘Well, I kept thinking about it and trying to picture the word, and then I thought... cos sometimes in year 7 we used to do the PowerPoint things where there’d be a picture and then the word, and that’s really helpful, so I was just thinking about that in my head, and I saw the snow and the word. So I said snow.’

By Time 9, Helen was reporting only the occasional word appearing in her mind’s eye, and this had disappeared entirely at Time 11. But at Time 16 she volunteered that the phenomenon had returned: in discussion about trying to puzzle out the word *vrai* (‘true):

‘But I couldn’t see it in my head, so I was like oh.’

So most of the time when you’re listening are you seeing things in your head?

‘Well, not if it’s automatic. If it’s not automatic, I try to, If I don’t know what the word is, I’ll try to think about how you’d say it. Like if you don’t know a word in English, you’d try to think about how you spell it.’

Do you force yourself to do that? Is it a conscious decision, or does it just pop up?

‘Not conscious.’

So it just pops up?

‘Yeah... I thought that was normal?... I just see the word. I can’t say how, if I’ve seen it before, then I can see it from there.’

⁵⁷ Helen in italics; me (the interviewer) in bold italics

And is it each word or do you see a whole sentence? Is it a word and then when the next word comes, do you see the two together?

'It depends how much I don't know or understand. It could just be a word, but it could be more.'

In fact, this 'self-subtitling' could be considered an element of the mental translation construct, as she identifies at Time 8 that when words come automatically to her, she does not 'see' them in the same way.

6.7.3. Phase 3: Growing confidence and proficiency

The third phase of listening identified by the change point analysis software began at Time 11 and continued until the end of the study. Recall that Time 11 (held in March 2020) was the last 'normal' time point before the pandemic, with all data collection taking place at home from Time 12 until Time 17 inclusive, and online lessons also taking place in the periods preceding Times 12, 13, 14 and 17. It was only a year later at Time 18 (March 2021) that pandemic restrictions again allowed me to collect data in the school setting.

In this phase, the correlations between listening proficiency and Directed Attention, and between listening proficiency and Person Knowledge, seemed to have fallen into step, with the moving correlation between listening proficiency and Mental Translation following a similar pattern two windows later. In other words, Directed Attention and Person-Knowledge can be said to be supportive of each other during this phase, and possibly facilitative of Mental Translation. By contrast, Problem-Solving and Directed Attention have become competitors: when Helen's listening proficiency is strongly

correlated with her Problem-Solving scores, it is negatively correlated with her Directed Attention, and vice versa.

In other words, during this phase we see complex dynamic interactions both within Helen's metacognitive subsystems, and between these subsystems and her wider listening proficiency. Even if the change point analysis software isolates this period as a single phase, Helen's listening continued to become more proficient.

The three-way relationship between Directed Attention, Person Knowledge and Mental Translation was apparent at the Time 11 interview, where not only did Helen report over-focus on specific elements of what she hears, driven by a determination to achieve, she also felt irritated when she noticed that her preoccupation with specific words led her to missing the rest of the input and therefore not achieving what she felt she should have done.

Helen's anxiety dropped considerably when lockdown began. She reported enjoying doing the data collection listening tests at home because they felt less competitive: during data collection in the school setting she had always been aware of what her peers were doing and comparing it to her own performance. This level of relaxation coincided with a moving away from a word-by-word translation, although not away from translation more generally. Instead, Helen set herself the challenge of translating full sentences. She told me that this was a deliberate attempt to challenge herself, and a curiosity as to whether this would increase her achievement. However, her stimulated recall at this time suggests that this attempt was not entirely successful; instead she discussed what can be interpreted as a Problem-Solving strategy interacting with Mental Translation and Directed Attention, where she is attempting to decipher the meaning of *avant* ('before') in the sentence *Je ne mange rien avant l'école* ('I eat nothing before school'):

You wrote I eat dot dot dot before school. Talk to me about that.

'I thought école was school, I just had it in the back of my mind. Then I heard I eat, like je mange and then 'before school' made sense if we were talking about breakfast. And also I thought I heard avant.'

Did you know the word avant?

'I didn't know it automatically, but I thought somewhere, I knew it was before.'

The moving correlations indicate that around Time 13 there was a very strong positive correlation between listening proficiency and Problem-Solving, and indeed this is evidenced not only by her accounts of her listening experience at that time, whereby she experienced moment-to-moment switches in the quality of her understanding, between automaticity of understanding, a semi-automaticity in which she is applying Problem-Solving strategies of use of context and co-text, and an effortful listening experience in which she took time to process what she had heard, sometimes at the expense of subsequent input.

'Now I just hear it and if I know it well, I hear it and basically just write it, I don't have to think about it. Other things, I don't know so well, I think about how it works in the context. If I really don't know, I'll think about what I actually heard. Then I think this could mean something. I think I translate some of the hard bits.'

In section 6.4.3 I discussed Helen's Time 14 translation of *chez ma cousine* as 'food/cuisine' from the linguistic perspective. However, this attempt might also be seen

as a Problem-Solving approach whereby whereby the metathesis of *chez ma* to *manger* ('eat') contributes to her choice of translation of 'cuisine' rather than 'cousin.

In the final phases of the study, Helen's automaticity grew further – and this can be seen as an automaticity not only of understanding, but also of approaches to understanding. At Time 18 she made a statement which challenges our traditional definition of language learning strategy as conscious approaches taken to facilitate understanding:

And then I just sort of like figure out the rest. But I'm not like 'ooh, what does that mean', I just figure it out somehow. I just take a second, and I'm like 'oh, that makes sense'. But it's quite hard to describe what I do in that second, because I don't really know myself.'

Her growing automaticity is also apparent in the experience of listening. Of the three levels of experience of listening that she had identified at Time 13 – those of automaticity, semi-automaticity and effortful listening, by Time 19 she reported that the effortful listening stage had largely passed, and this could be concluded in our final interview at Time 20, where she reported:

'Now when I hear something in French, I'm not completely overwhelmed... some of it's automatic just because I've heard it so many times. It's just there. And there are other parts of it, [where] I use the context a lot to, like, piece it together.'

6.7.4. Metacognitive subsystems: Concluding remarks

- Helen reported a drop in anxiety throughout the study, moving from very anxious to not particularly anxious. She also translated less and less in her head over time.

- Both these changes appear to align with the change points found at Time 7 and Time 11.
- Her Directed Attention improved during the study.
- Directed Attention and Problem-Solving seemed to compete after Time 11.
- Her ‘self-subtitling’ approach to listening is distinctive.
- Automaticity both in understanding and in problem-solving grew significantly in the final phase of the study.

6.8. Helen’s complex dynamics: conclusion

In examining a single case of Helen, it is clear that although her listening experiences bear some similarities with that of the mean findings explored in Findings 2, there are also unique elements to Helen’s complex dynamic system. The fluctuations in all subsystems are inevitably far more dramatic than will be seen with a sample size of 80. Even while out of the four case studies, Helen’s listening trajectory bears the closest resemblance to that of the mean, we saw very different data coming from the moving correlations, suggesting that each individual is experiencing their own complex dynamic system and for each individual different subsystems are playing different roles at any given time.

The lower subsystem scores relative to Helen’s overall listening echo what has been found qualitatively: that Helen’s ultimately very high listening proficiency is rooted in canny use of strategy and determination to succeed at least as much as it is in lexical and syntactic proficiency and progress. In examining the experiences of the other case studies, we can now therefore begin to probe the question of whether it is possible to reach the same end point in proficiency via a range of different routes.

7. Findings 4: Theo

This is the second of the four case studies. This chapter explores the experience of ‘Theo’: his listening processes and development will be uncovered through both his errors and the changing relationships between his subsystems. Although Theo and Helen both performed in the top quartile for listening proficiency, this chapter will demonstrate how Theo’s complex dynamic systems result in a different trajectory.

7.1. Theo: Background information

Theo lives with his mother, father and younger brother in the city. He was a ‘higher-attaining’ boy. All participants were born within a year of each other, but Theo is one of the oldest. Research from the UK school system suggests that older children in a cohort achieve more highly than younger ones (Balestra, Eugster, & Liebert, 2020; J. F. Bell & Daniels, 1990). Although I chose a male high prior attainer at random for my case study when all the participants began secondary school, Theo achieved the single highest score in the listening proficiency tests.

School data reported that English was the first language of the home with no mention of additional languages. Yet, Theo told me that his mother spoke Swahili and she used it with him on occasions: he understood it, but he rarely produced it himself. He had done a small amount of French at primary school, but said that this had consisted almost entirely of vocabulary items and that he had not been given the opportunity to form sentences. Throughout the data collection period, he had a little out-of-class exposure to French: he overheard tourists speak French on a bus, he spent some time attempting to read the *Le Petit Nicolas*, a French children’s book, and he listened to the Belgian musician Stromae.

During the first covid lockdown he exchanged WhatsApp messages with a French boy on a weekly basis.

The theme running through our interviews was one of ambition and a confidence that his determination to succeed would come to fruition. He talked regularly of enjoying learning, the value in learning French and the importance of academic subjects – at Time 1 he remarked:

‘When I see a lesson I like, I’ll prefer to come to school than when I’ve got like PE and art a lot of the day.’

At the beginning of the study he identified French as his favourite subject, defining it ‘like a code you have to crack,’ Although his love of French dipped when he changed teachers in year 8 (from Time 7 to Time 13), he continued to relish the intellectual challenge that his learning was providing. In regards to the Time 9 test battery, he said,

‘It was hard but kind of fun. It’s kind of fun when you don’t know something, but you can try and guess. It’s like you have to think more and you have to focus more.’

At the end of the study, reflecting on his three years at secondary school, he said:

‘The learning’s gotten harder, which is good, because it’s less boring.’

In the final interviews, from Time 16 to Time 20, Theo was keen to explore listening pedagogy with me and in response to such questions as ‘is there anything else you’d like to add,’ would reflect on the difference between the design of my paused translation tests and typical classroom listening tasks. While he told me at Time 16 that,

‘Listening for an answer is easier. I can hone in on one thing. I listen out for that and zone out for a bit until I hear the word,’ he also recognised that *‘listening for the full message is more interesting.’*

By Time 18 he was more scathing about the Comprehension Approach, saying normal listening tasks in lessons are,

‘All easy ‘cos you literally just have to know what the verb is. For example, if it’s saying, “Last year, who went cycling,” you literally just have to look for the word “went cycling” and then it’s fine. Probably not even “went”. Just “cycling”.’

At Time 19 he had reflected that he had not been taught,

‘How to unpick what you hear. There’s no method to listening; you should just know what the word means.’

But despite his dissatisfaction with listening pedagogy, he appeared to be happy throughout his time at school, and despite the two covid school closures, he developed a close friendship group of equally academic and ambitious peers. Theo also played the clarinet at a high level, participating in many school concerts, and represented the county at swimming.

7.2. Theo’s listening proficiency and its variability

As with all the previous findings, Theo’s listening proficiency is set on a scale of 0 to 100, with 100 roughly equating to a CEFR grade of B1, or a GCSE grade 4 to 5. In Theo’s case there was a very strong correlation of 0.79 between time point and listening proficiency score, indicating that good progress was made during the study.

7.2.1. Variability in listening proficiency scores

Table 46 and Figure 71 give Theo’s raw listening scores. Theo’s listening scores were highly variable. It is notable that Theo’s initial score at Time 1 was only slightly above the cohort mean, but after that he made rapid and significant progress, with scores for listening proficiency between 20 and 48 points above the mean.

Table 46: Theo’s listening progress and the cohort mean

Time	Cohort mean	Theo
1	36.40	39.40
2	40.06	53.45
3	40.26	56.14
4	44.14	61.12
5	41.85	67.62
6	41.59	58.64
7	47.42	72.68
8	41.99	84.16
9	44.62	70.35
10	47.45	68.60
11	48.87	70.33
12	49.07	80.43
13	49.30	72.17
14	47.03	68.83
15	51.66	100.00
16	51.51	87.24
17	50.89	74.96
18	49.36	83.15
19	52.86	98.49
20	50.83	77.70

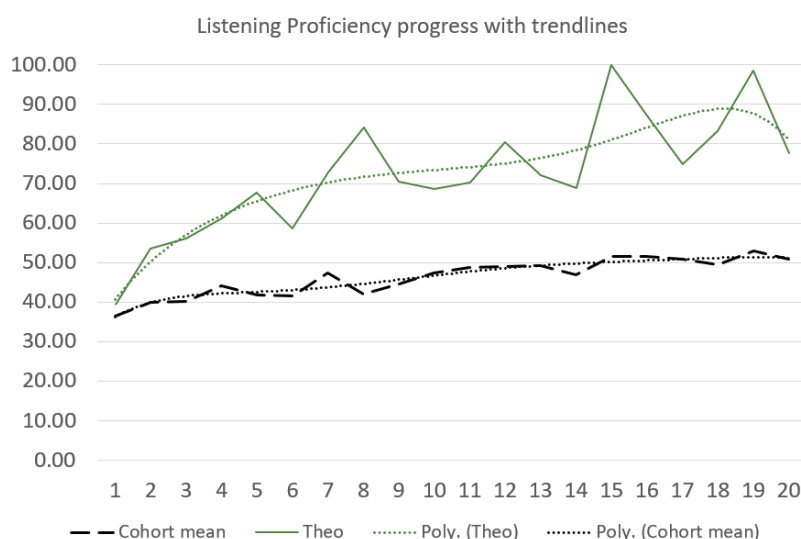


Figure 71: Theo’s listening progress, with trendline

Theo talked about the lack of listening tasks during the online learning of the two lockdowns: at Time 12 (May 2020) he told me that ‘*I have to adjust to listening in French*’, and at Time 13 (July 2020) – having done no listening at all since the Time 12 test battery – that listening ‘*took some getting used to*’. After six weeks back at school, at Time 15, when Theo excelled in his listening proficiency, he talked of, ‘*listening to lots of French and having that objective again of thinking about it properly.*’

When viewed through the lens of the moving min-max graph (Figure 72), we see rapid progress initially made in Theo’s first year of French study followed by a short period of

relative stability between Time 8 and Time 12. This timeframe equates to the beginning of the second year of study until the first covid lockdown, which began during Time 11 data collection. Theo’s highest score – indeed, the highest observed score in the study – was achieved at Time 15, on the return to school in September 2020. The end of the moving min-max graphs hints at a possible convergence, suggesting that a new state of stability at a higher proficiency level might have been being reached. Recall that the instability of the bandwidth – the distance between the minimum and the maximum score within a given window – indicates a state of flux due to the wide range of proficiency. By contrast a narrower bandwidth by definition shows less change, or a period of relative stability of progress.

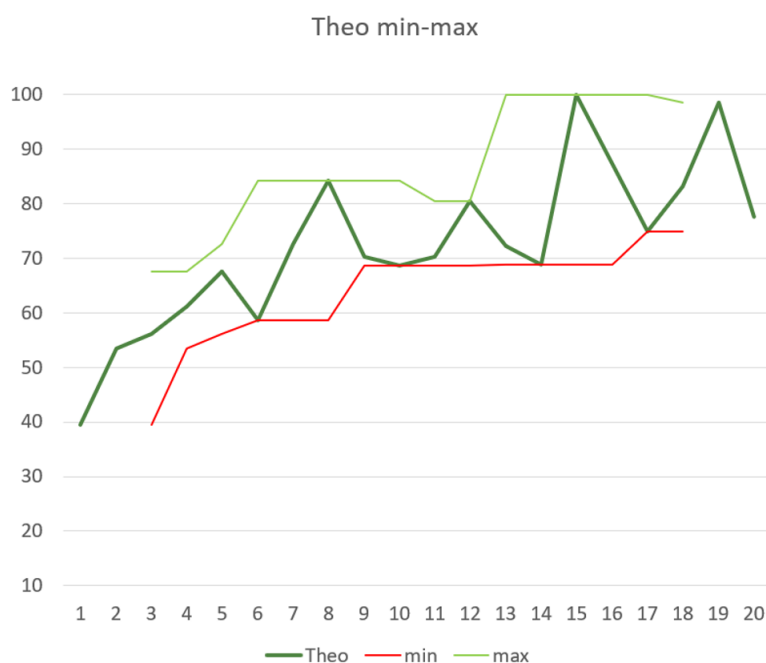


Figure 72: Moving min-max for Theo’s listening proficiency

The change point analysis software (Taylor, 2000) reveals a single change point in Theo’s trajectory, with the first phase comprising Times 1 to 6, which equates to year 7, and the second phase comprising the remainder of the study. This can be seen in Figure 73, and indicates a mean score in Theo’s first phase of the study of 56.06, moving to 79.22 in his

second phase. For a reminder of the procedures involved in the change point analysis, refer to Methodology section 3.6.3.

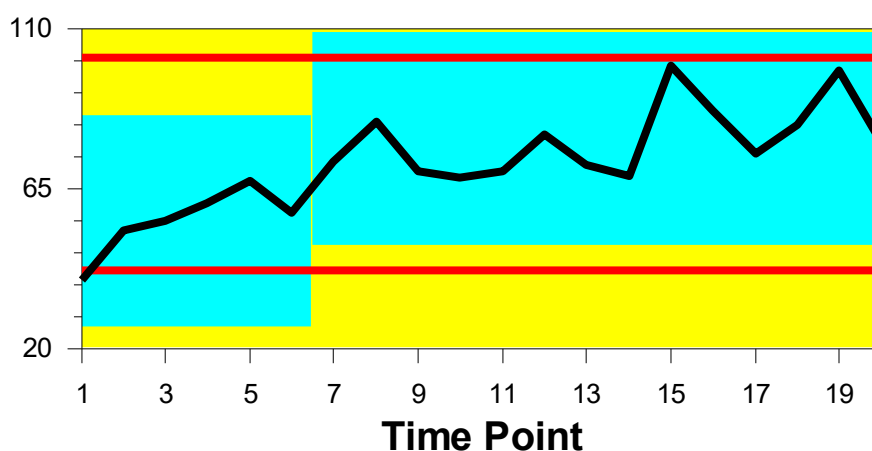


Figure 73: Change point plot for Theo

7.3. Theo's linguistic subsystems, and their development and relationships

7.3.1. AuralLex scores

Theo's vocabulary scored as measured by the AuralLex test, compared with the mean, are given in Figure 74. His vocabulary size began higher than the mean and with the exception of Time 8 and Time 10, exceeded the mean throughout the study, although large fluctuations are apparent. Recall that four tests of equal difficulty were rotated throughout the study – these are colour coded in the x axis. In other words, the time points shaded in red were the same

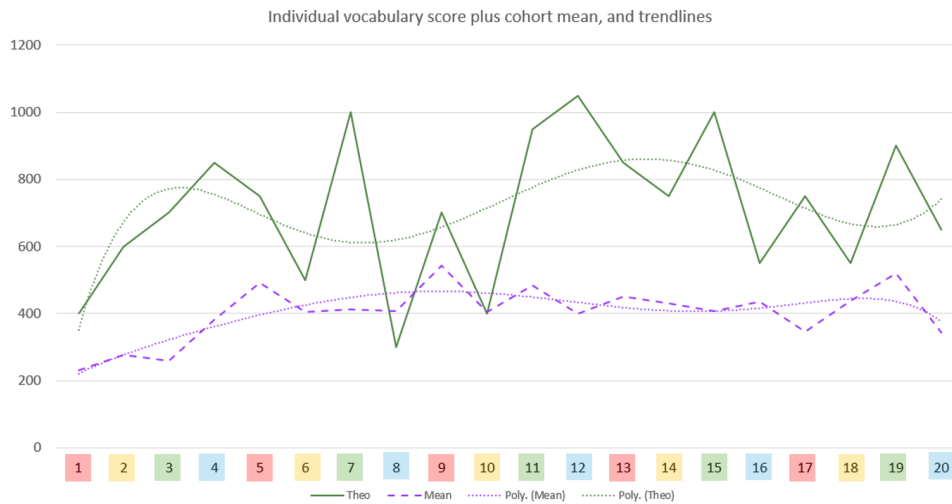


Figure 74: Theo's AuralLex scores compared with the mean

test, as were numbers shaded in orange, and so on. This demonstrates that the fluctuations cannot be said to be due to the specific tests, but other factors were at play. Theo's highest vocabulary size was measured at Time 12 at 1050. (For context, we saw in Findings 2 that the highest vocabulary size measured in the entire study was 1250 at Time 8 and the highest mean was 420 measured at Time 12).

Theo's vocabulary score as moving min-max is given in Figure 75. The very wide bandwidth (distance between minimal and maximum scores) between Time 6 and Time 12 is indicative of the fluctuations between vocabulary size measured at different time points, showing a system that is unstable. In other words, Theo's vocabulary knowledge was unpredictable during this time. The increase in moving minimum score from Time 14 onwards suggests that more stability was beginning to show within the system.

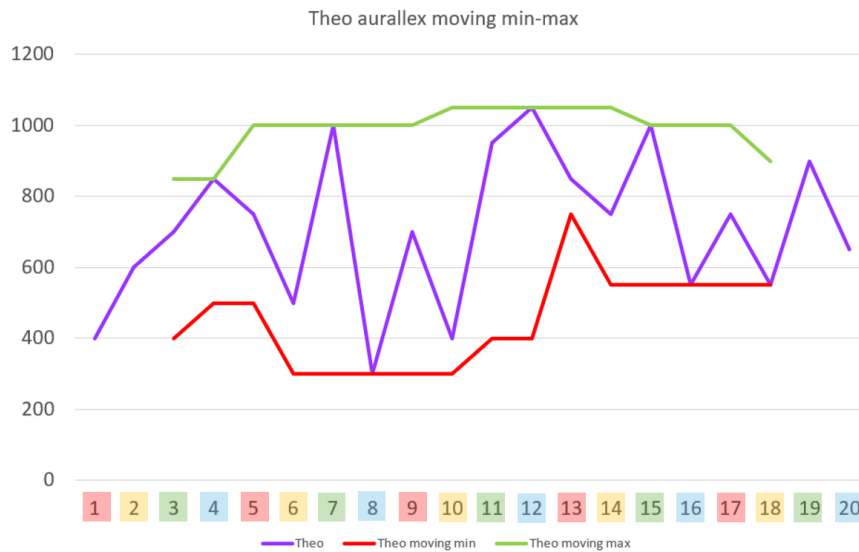


Figure 75: Theo's moving min-max vocabulary size

Theo hinted at a possible explanation for the very large fluctuations in his AuralLex scores in our Time 2 interview. Recall that participants would hear a word and would then be given the option to select either 'I know or can use this word' or 'I do not know this word'. Theo reflected deeply on the implications of these two options, saying,

'If I hear it and I'm like, "I've heard that word somewhere before, but I can't put my finger on it," I don't know whether I should click "I know it" or "I don't know it". Because it could be a cognate and I could just be thinking, "well it sounds like [a word] in English".'

7.3.2. AuraGram scores

Theo's grammatical development as measured by the AuraGram is shown in Figure 76. Recall that this was measured via six aural grammaticality judgement tests of equal difficulty; they are colour coded in the x axis to indicate test repetition. Theo's score was consistently above average, although again demonstrated fluctuations which cannot be accounted for by the specific test. Theo's highest AuraGram score was 12/20, at Time 7

and Time 17 (and at both these time points 12/20 was the highest score); the highest score achieved within the cohort was 15/20 at Time 15.

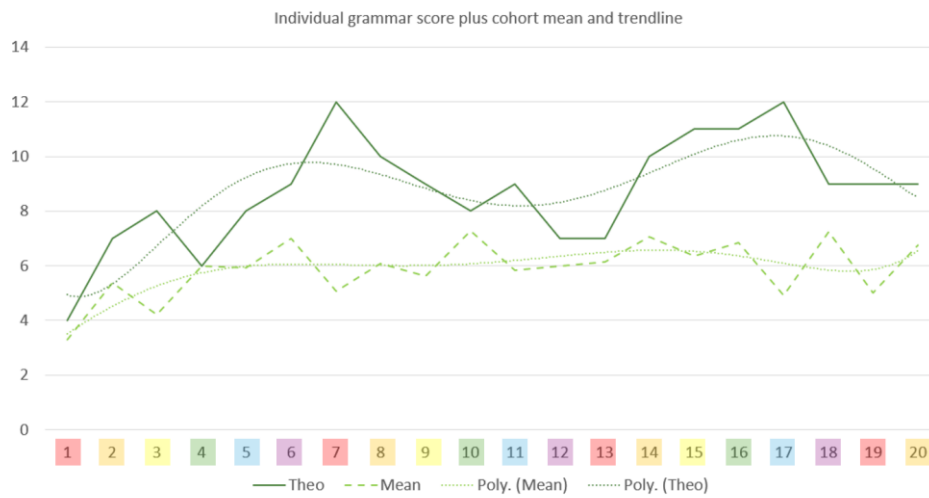


Figure 76: Theo's grammar scores plus cohort mean

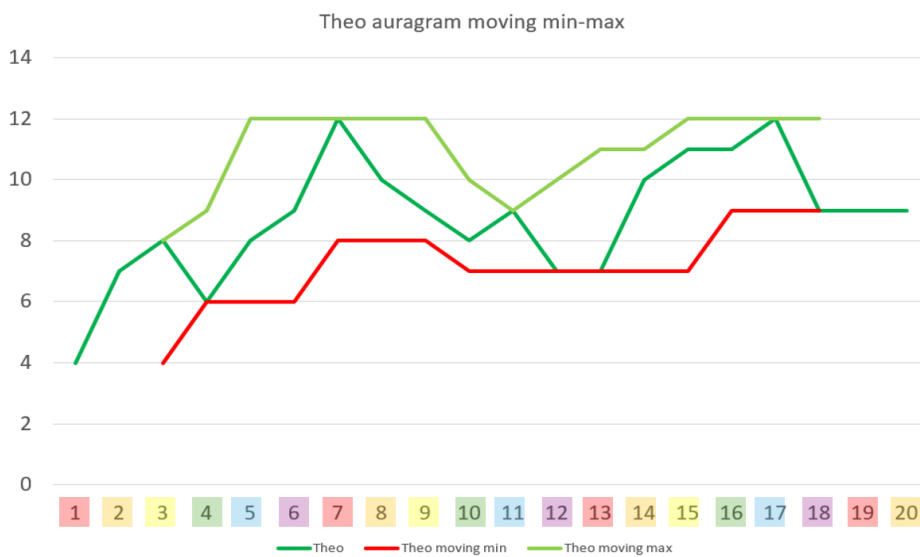


Figure 77: Theo's moving min-max grammar scores

Theo's moving min-max scores are shown in Figure 77, highlighting fluctuations in bandwidth and illustrating erratic performance throughout the data collection period. In our Time 18 interview, Theo said,

‘If there was a chart to show how much grammar I’ve learned, I feel like it spikes and then stops and then spikes and then stops, and then spikes again. But I still think it’s pretty poor, just cos of how difficult French grammar is.’

Theo shows insight into the uneven trajectory of his grammar knowledge in this way.

7.3.3. Developing relationship between linguistic subsystems

In order to address the relative roles of lexical and grammatical knowledge in Theo’s developmental trajectory of listening, Figure 78 places Theo’s vocabulary, grammar and listening scores on the same graph, normalised on a scale of 0 to 1. As stated in section 5.2.1 of Chapter 5, the 0.5 line represents a vocabulary size of 600 words. In grammar terms, the maximum observed score was 15/20, so the 0.45 line represents a score of 7/20.

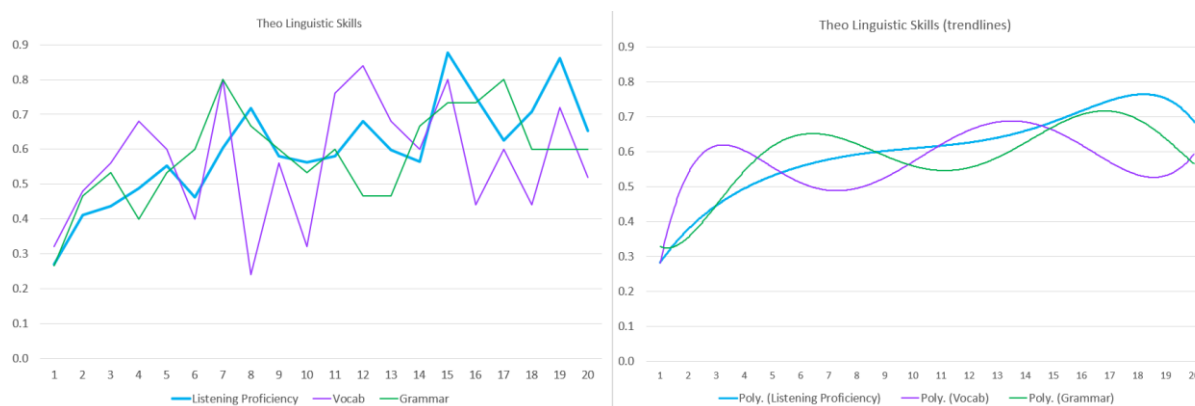


Figure 78: Theo’s vocabulary, grammar and listening scores on the same graph

The progress of vocabulary and grammar knowledge is shown in Figure 79, with the listening proficiency line removed. Examined through this lens it appears that, for Theo, when grammar knowledge was increasing, vocabulary was dropping, and vice versa.

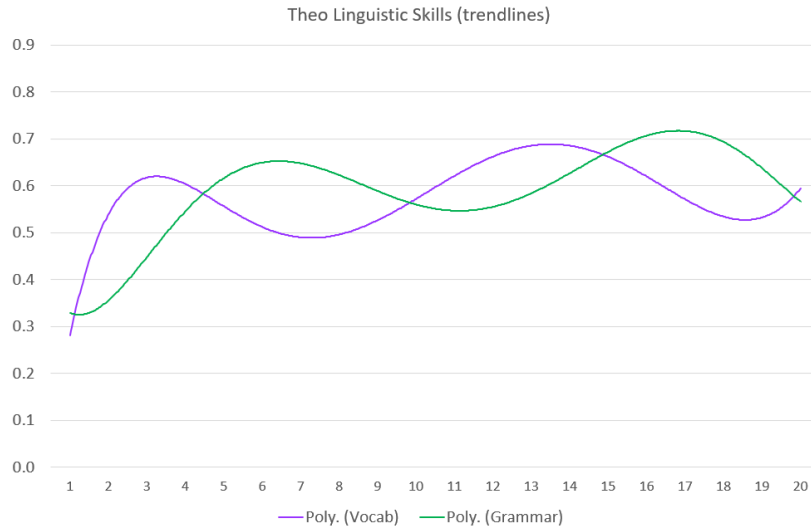


Figure 79: Theo's trendlines for vocabulary and grammar knowledge

Moving correlations between listening proficiency and both vocabulary and grammar are given in Figure 80. They indicate a highly variable relationship.

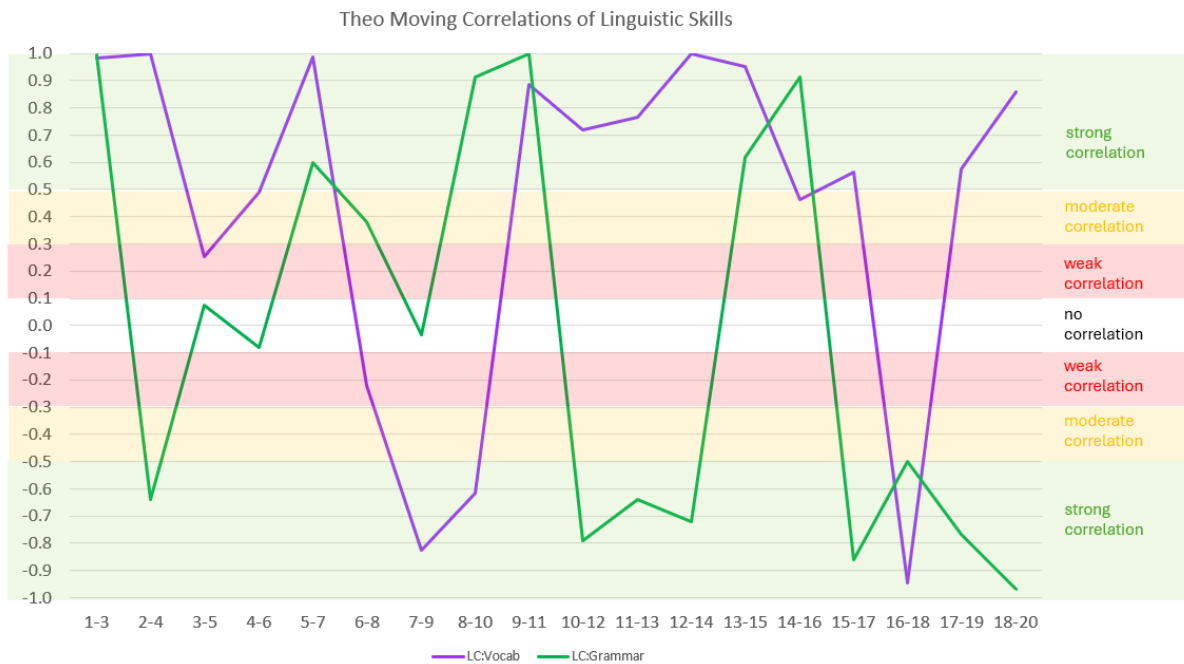


Figure 80: Moving correlations of vocabulary and grammar for Theo

The strong or moderate moving correlation between vocabulary and listening proficiency ended after the T5-7 window, in fact becoming a strongly negative relationship for a

short period before returning to another period of positivity. This period of strong positive correlation was interrupted solely by the T16-18 window where again the relationship becomes strongly negative (i.e. here listening proficiency increased while vocabulary size appeared to fall).

Theo's moving correlation scores between listening proficiency and grammar (green) were much more erratic. When comparing the trajectory of the vocabulary / listening relationship with that of the grammar / listening, there was some indication of a competitive relationship between grammar and vocabulary between the T9-11 window and the T13-15 window, in that vocabulary is strongly positive while grammar is strongly negative. This moved into a brief 'precursor' period whereby the strongly negative correlation between listening and grammar at T15-17 preceded a similar dip in the relationship between listening and vocabulary at T16-18. Yet at the end of the data collection period, the two subsystems appeared to have reverted to their previous competitive relationship, as the relationship between listening and vocabulary became strongly positive again and the relationship between listening and grammar remained strongly negative.

7.4. Theo's errors and listening processes

This section will draw on interview data and test results to explore Theo's errors and listening processes. In Chapter 4, I separated the types of errors into whole-track segmentation difficulties, vocabulary difficulties which appeared to be driven by either phonological challenges or lexical gaps, and grammatical errors. We will see in this section that Theo's errors are largely rooted in lexical gaps; he appears quite unusual within the cohort in that the phonological nature of the tasks posed minimal problems for

him. For example, within the entire dataset he only twice complained that the input was too fast – once at Time 2, and once at Time 11.

7.4.1. A listener unfazed by the French speech stream

Minimal pairs were shown to be a struggle for the full cohort, as shown in section 4.3.3. Theo was an exception, as only four cases of this difficulty occurred within the entire dataset: at Time 3 he confused *vélo* /velo/ and *volley* /vole/; at Time 9 *soir* /swaʁ/ (‘evening’) was translated as ‘sister’ /sœʁ/ (*soeur*) and at Time 18, *on est allés* /ɔ̃n_ɛt_ale/ was translated as ‘In Italy’ (/ɔ̃n_itali /). Finally, in response to *je me suis bien amusé* /ʒə mə syi bjɛ̃n amyse/ he told me, ‘I just heard “amusant” or “amusé”’. In other words, although Theo had some listening difficulties that were typical of the cohort at large, he did not appear to struggle to the same extent.

And indeed, where Theo’s listening difficulties were phonological, they almost always respected word boundaries. (*On est allés* / *en Italie*, above, is a notable exception.) In section 4.3.1 of Chapter 4 I discussed the sample’s translation of *bouteille d’eau* as ‘potato’ within the context of segmentation difficulties. Theo did not translate this correctly, but his mistranslation of *bouteille* /butɛj/ as ‘butter’ did not attempt to run several words together, as many of his fellow students were inclined to do.

7.4.2. Puzzling it out with vocabulary at the core

In lieu of challenges driven by the phonological nature of the listening experience, then, Theo’s difficulties often appeared to him to be rooted in a simple lack of vocabulary. Frequently during stimulated recall, we would discuss a perfect translation of a track, and his verdict would be that, ‘I just knew it.’ From the outset, he would use context alongside vocabulary to lead him to a correct translation. His thought processes were well elucidated at Time 2 when he was asked to translate *je ne joue pas d’instrument mais je*

suis chanteur dans un groupe ('I do not play an instrument but I am a singer in a group').

He said:

They said ne and then a blank and then pas, which means it's negative, which probably means like either they don't play it or they don't like. And then they said ma which means 'but'. And then they said je chante dans which I think means 'but I sing'. So he doesn't do or doesn't like something, but then he sings, so maybe he doesn't... and then at the end it said something like group, so maybe he doesn't like... no... he doesn't play an instrument, but then he likes to sing in the group.

Furthermore, where Theo struggled, he told me that he would be inclined to use context (more on this in section 7.7), but if this approach let him down, he would miss out the word. An example was seen at Time 12, where, in *mais j'adore mettre du miel sur ma tartine* ('But I love to put some honey on my bread and butter'), Theo just omitted the word *mettre* ('to put'). In both these cases, he therefore reached a serviceable if not entirely accurate translation.

This inclination to omit was also seen at Time 19, where in response to *je peux l'utiliser jusqu'à neuf heures du soir* ('I can use it until 9pm'), he wrote '*I can use it in the evenings*'. When pushed to explain his reasoning, he told me:

'Je peux – so, "I can". And then I heard utiliser and I just thought OK, that means "to use", so, "I can use it". And then I heard le soir at the end, but I don't... I heard other words in between but I must have disregarded them because I don't know what they meant.'

In other words, Theo used context to his advantage where he could, but was also able to recognise occasions in which accurate translations could not be reached with the help of

context, and omitted them. As such it is likely that for Theo, a moment-to-moment complex dynamic system was arising in his listening in which vocabulary and context worked together, leading him either to the right answer, or to one which was ‘good enough’.

We can see that Theo’s approach to omitting words was unusual when comparing his experience with the whole cohort. In Section 4.5.1 of Chapter 4 I reported on the differences between content and function words in this regard. In Table 47 I reproduce Table 38 in percentages to enable comparison with Theo’s data:

Table 47: Missing and wrong function and content words: Comparing Theo with the full cohort

	Missing		Wrong		Total	
	Theo	Cohort	Theo	Cohort	Theo	Cohort
Content word	38.9	20.1	38.9	57.9	77.8	78.1
Function word	14.8	19.3	7.4	2.6	22.2	21.9
Total	53.7	39.5	46.3	60.5	100.0	100.0

Table 47 demonstrates that although the total proportions coded to function word difficulties and content word difficulties were very similar for Theo’s case and for the full sample, the breakdowns are quite different. Theo was much more likely than average to omit a content word and much less likely than average to provide an incorrect translation of a content word. By contrast, he was much more likely than average to offer up a wrong function word. These data suggest that Theo’s own explanation that he omitted what he did not know was not entirely accurate. Perhaps more accurately: he omitted what he *knew* he did not know. We have seen that he had excellent vocabulary,

but the data suggest that his confidence in correctly identifying function words was greater than that of the full cohort.

7.4.3. A preoccupation with verbs yet overlooking morphological features

In our interviews, Theo regularly referenced the importance of verbs for his meaning-building, and it became a mainstay of his approach to listening, telling me at Time 10 that it was *'the most important part of the sentence'*. Indeed, on the one occasion when he reflected on the role of his additional language of Swahili (at Time 16), it was in the context of verbs:

'Because Mum tells me, go do this or shut this. I'm normally listening for the verb and then the object I need to do the verb to. I think that helps with my honing of a particular part of a sentence and knowing what to listen for.'

At Time 17, the cohort had returned to lockdown learning and, out of practice in processing auditory input, Theo said he was *'reverting to thinking about the verb... and then listening around it.'* He also recognised that he was struggling to hear the difference between tenses. Theo saw the role of verbs as central to his learning:

'I'd say the main thing is learning how to change verbs, learn new verbs, and then just adding words around the verbs.'

In fact, when I asked him to reflect on the previous three years of French, he even joked that in lieu of the word 'French' on his timetable, it should simply read 'verbs'.

Yet despite this very raised awareness of the importance of verbs, and a determination to focus on them when listening, Theo often struggled with recognising the morphological features connected to different tenses in French. In other words, borrowing the concept

of the distinction between content and function words, for him the focus of the verb was its content (meaning) rather than its morphological function-features (tense, person).

Nonetheless, a definite developmental trajectory was apparent with Theo's translation of verbs. In test D25, recognition of the near future tense was tested in several of the tracks, at Times 9, 13 and 18. Theo translated *je vais regarder la télé* ('I'm going to watch TV') as I watch TV at Time 9, but dealt correctly with all tense issues in subsequent iterations of the test, even though the near future tense was not a specific focus of lessons at Times 13 or 18. A similar pattern was visible with the translation of *je prends un photo* ('I take a photo')⁵⁸, which started as 'I took the photo' at Time 4 before reverting to a correct translation at Times 7, 12 and 15. The fact that Theo was correctly identifying tenses despite their not being a feature of lessons at the time demonstrates further mastery of the language.

Test D8, which appeared at Times 8, 12, 16 and 20 tested recognition of a morphological feature of two verbs with both *j'apprends à jouer* ('I'm learning to play'), in which participants needed to identify the /e/ ending as an infinitive and *je reste à la maison* ('I'm staying at home'), where they needed to notice the absence of the infinitive /e/ ending and recognise a conjugated verb in the present tense. Theo's translations are given in Table 48. It is noteworthy that at Time 8, Theo recognised the /e/ ending in *jouer* which indicated the infinitive yet did not recognise its absence in *je reste à la maison*, which was the very next track. In subsequent iterations of the test, the mistranslation of *je reste* was resolved, but Theo continued to struggle with *j'apprends à jouer*, all the while remaining faithful nonetheless to the presence of the infinitive.

⁵⁸ Test E26 track 1, tested at T17, T12, T15

Table 48: Trajectory of Theo's translations of *J'apprends à jouer* and *Je reste à la maison*

	J'apprends à jouer	Je reste à la maison
Time 8	I to play	I like to stay
Time 12	[no attempt made]	I stay
Time 16	I take to play	I stay
Time 20	I have taken to play	I stay

His Time 16 mistranslation confused *j'apprends* /ʒapʁɑ̃/ with *je prends* (/ʒə pʁɑ̃/ 'I take'), and at Time 20, his rendering in the past tense is probably explained by a recognition that the initial sound is not /ʒə/, which would indicate a present tense, but /ʒa/. (Correctly, in the past tense this would have begun with *j'ai* /ʒɛ/.)

In fact, this last example could also be called on to illustrate that, throughout the study, Theo struggled with the aural recognition of a range of morphological features. There were 12 incidences of failure to detect plurals within Theo's dataset, most notably in test D8 (as discussed in section 4.5.2 of Chapter 4 and 6.4.5 of Chapter 6 – Case Study Helen), which persisted until the end of the study. Theo acknowledged that in order to recognise tense markers he listened out for confirmation in the form of time phrases: quizzing him at Time 18 on his correct translation of *l'année dernière je suis allé* ('Last year I went'), he said '*l'année dernière, that's what set it for me. Because that means last year*'. Where time phrases are commonplace to accompany specific tenses, within the dataset there were rarely equivalent lexical 'boosters' to help a listener detect singularity or plurality. This might indicate why apparent mastery of tenses came sooner to Theo than the mastery of noun quantity.

7.5. Summary of findings for the linguistic systems

For Theo we have seen:

- A steady increase in listening proficiency throughout the data collection period, with a single change point statistically identified at Time 7.
- An overall growth in vocabulary size well above the cohort average, but characterised by many large fluctuations from time point to time point.
- Grammar scores which also sat higher than the cohort average, but which also showed marked fluctuations during the study.
- Moving correlations and trendlines which suggest that Theo could progress either with vocabulary or with grammar at any one time, but not with both
- Qualitative data which indicate that Theo's key difficulties with listening were lexical or morphological, but rarely segmentation-related.
- An understanding driven by the identification and analysis of verbs, although the ability to identify morphological features was still unreliable at the end of the study.

7.6. Theo's metacognitive subsystems, their development and relationships

7.6.1. Metacognitive scores

Theo's scores for the metacognitive subsystems are given in Figure 81, relative to the mean, with moving min-max scores in Figure 82. Recall that the statements were rated on a six-point Likert scale ranging from 'strongly agree' to 'strongly disagree', where 'strongly agree' would equate to a score of 1.0. The statements on the MALQ are in appendix 15.1.1.

Figure 81 shows slightly fluctuating Person Knowledge scores (which relate to a construct of anxiety) beginning well below average but rising from Time 7 (the point at which the Change Point Analyser indicated a new phase) to Time 11, where it peaked in line with the cohort average at a score equivalent to ‘agree a bit’. Theo’s Person Knowledge score then dropped again but showed another rise from Time 15 to the end of the study. By contrast, Theo’s scores for Directed Attention, Mental Translation and Problem-Solving were almost always above the cohort average, with the exception of a brief dip in Directed Attention around Times 5 to 7, and a small dip in Mental Translation around Time 15. Theo’s Problem-Solving scores showed a gentle decrease with time, whereby they began close to ‘strongly agree’ and ended with ‘agree quite a lot’.



Figure 81: Theo’s metacognitive scores relative to the mean

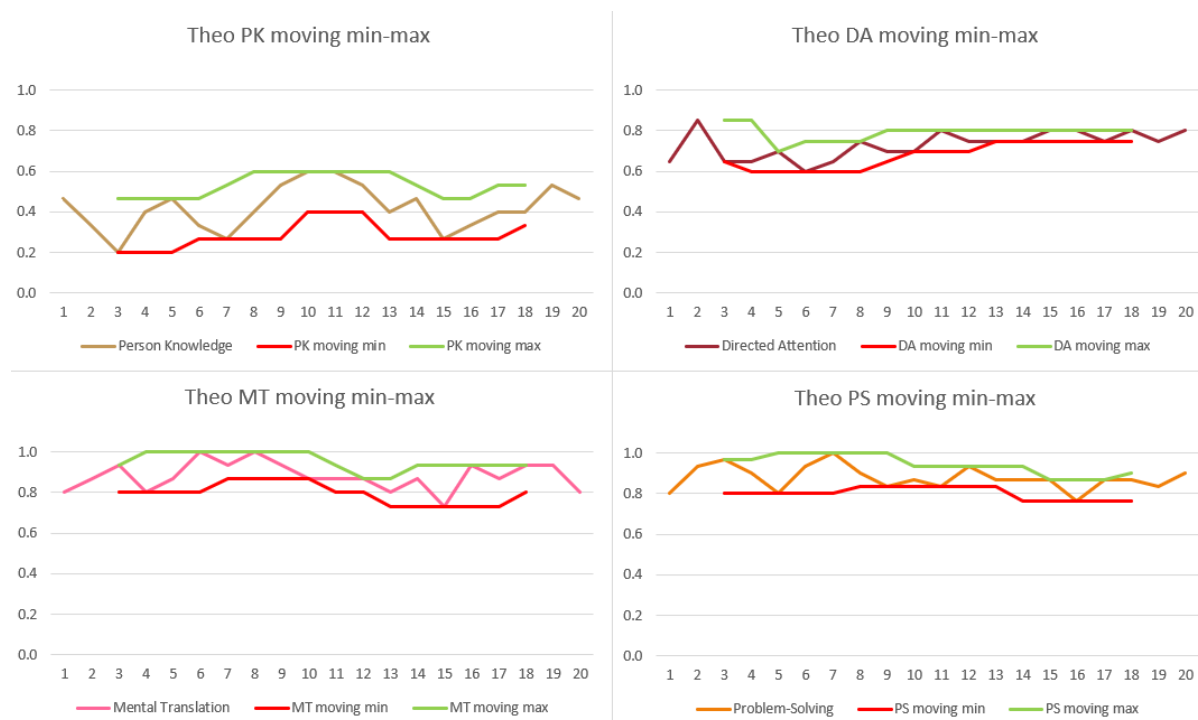


Figure 82: Theo's metacognitive scores through a moving min-max lens

The moving min-max graphs in Figure 82 point to a development in Person Knowledge after Time 6. However the graphs do not indicate any other metacognitive events taking place at the same time. The narrowing of the Directed Attention moving min-max suggests that this element of Theo's listening strategy became increasingly stable over time. By contrast his Mental Translation continued to waver. Like his Directed Attention scores, Theo's Problem-Solving scored showed signs of reaching stability between Time 10 and Time 13, but then fluctuation restarted.

7.6.2. Developing relationships between metacognitive scores

In section 5.3.2 of Findings 2, I showed that on average, trajectories of Mental Translation (pink) and Problem-Solving (orange) followed each other, and that the same was true of Directed Attention (burgundy) and Person Knowledge (fawn). But in Findings 3.1 I showed that Helen aligned Directed Attention with Problem-Solving

instead, and Mental Translation and Person Knowledge. Figure 83, showing Theo’s metacognitive trajectory on a single grid, indicates a different pattern again.

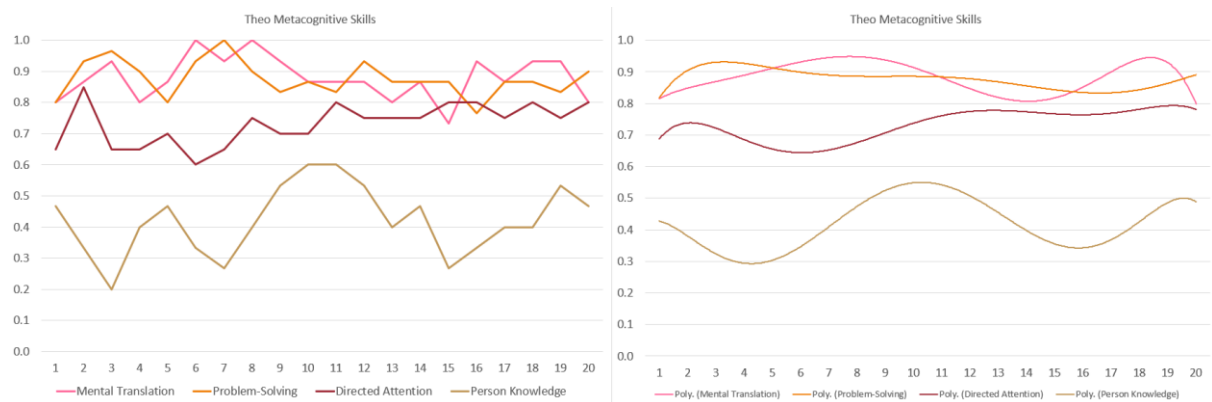


Figure 83: Theo’s metacognitive trajectory

For Theo, then, there was a Mental Translation / Problem-Solving relationship, just as there was for the full cohort, but Directed Attention and Person Knowledge do not appear to align, nor did any relationship emerge between the two.

Figure 84 illustrates the moving correlations between Theo’s listening proficiency and each of the four metacognitive subsystems. The moving correlation between Person Knowledge and listening proficiency fluctuated between strongly positive and strongly negative relationships over time. There was a strongly positive relationship between the two constructs between the Time 2-4 window and the Time 4-6 window, but we see a strongly negative relationship from the Time 8-10 window, which coincided with the Change Point, and also a period of less progress and higher levels of anxiety. Despite a brief period of strong positivity in the Time 12-14 window (the first lockdown), on return to school the relationship returned to strongly negative. In other words, higher Person Knowledge (anxiety) scores were correlated with less listening progress. It is worth remarking at this juncture, though, that we have already seen that Theo’s Person Knowledge scores were never particularly high (moving from an equivalent of ‘not very

anxious’ to ‘slightly anxious’), so although the moving correlations are shown to be strong, the underlying fluctuations were small.

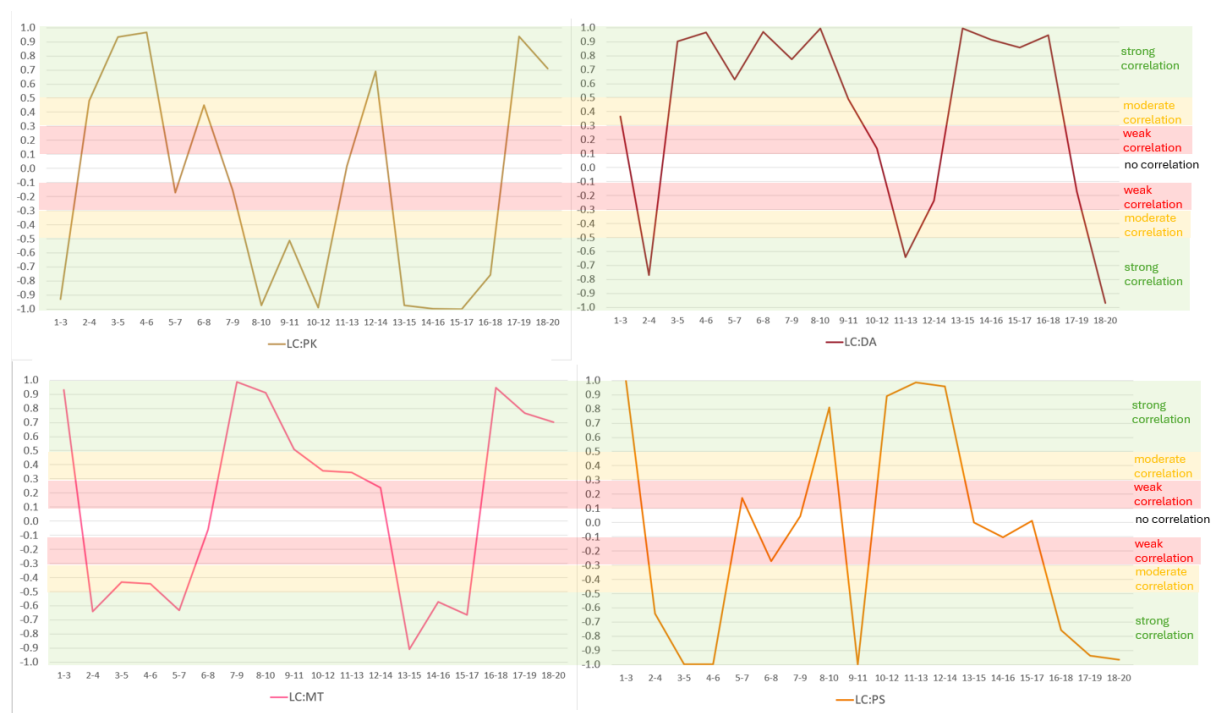


Figure 84: Theo’s moving correlations between listening proficiency and the four metacognitive systems

By contrast with Person Knowledge, the moving correlations between listening proficiency and Directed Attention largely show a strong positive correlation, with the exception of the Time 2-4 window, the Time 11-13 window (the first lockdown) and the very end of the study at Time 18-20.

Given the close relationship shown in Figure 83 between Mental Translation and Problem-Solving, I have also placed the moving correlations between listening proficiency and these two constructs on a single graph, which can be seen in Figure 85.

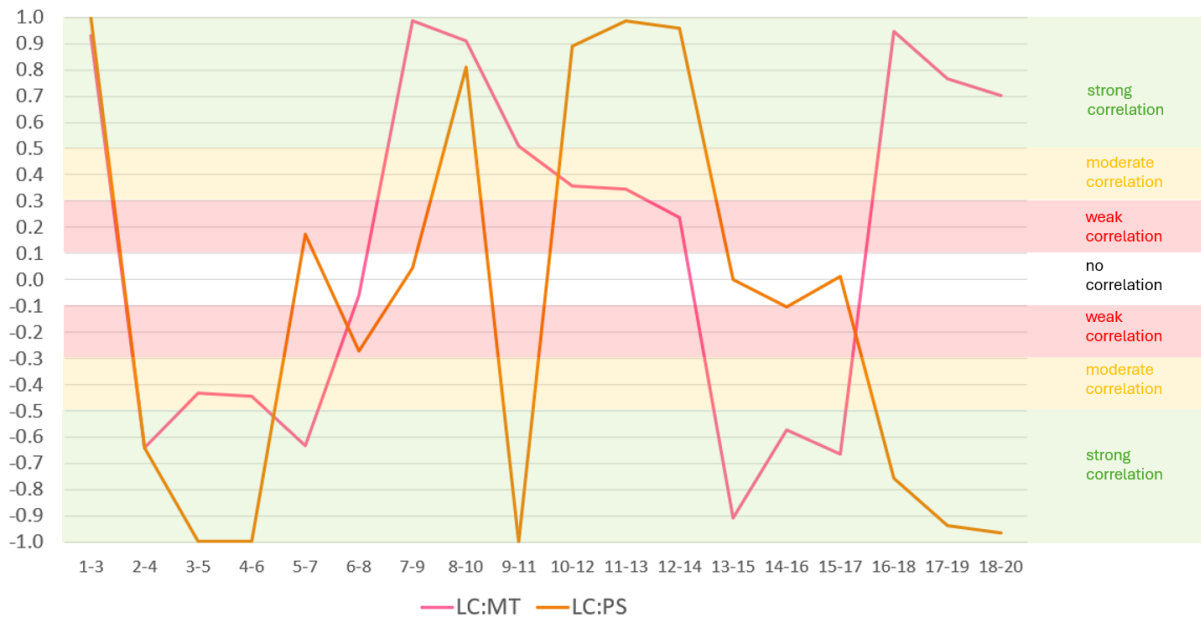


Figure 85: Theo's moving correlations between listening proficiency and both Mental Translation and Problem-Solving

Problem-Solving addresses the strategic use of context and co-text; and Mental Translation asks participants about the extent to which they translate as they listen. Figure 85 suggests that in the first half of the study, Theo's Mental Translation and Problem-Solving systems were broadly aligned and their relationships with listening proficiency were similar. However in the second half – shortly after the Change Point identified at Time 7 – the two systems appear to have moved into a more competitive relationship, whereby strong positive correlations between listening proficiency and one of the subsystems was matched by a strong negative correlation between listening proficiency and the other. In other words, perhaps for Theo's listening strategy, Theo used context to help him with the mental translation, but later, as his vocabulary size developed (see Figure 74), he became less reliant on context and co-text to reinforce his translation hypotheses.

7.7. Theo's strategic journey

This section will explore the qualitative data collected through the semi-structured interviews which took place at each time point, to build an image of Theo's developmental trajectory in metacognitive listening skills.

7.7.1. A listener who relishes challenge and expects to do well

Figure 81 illustrated that Theo's Person Knowledge⁵⁹ scores averaged at 'disagree a bit' during the first year of the study (to Time 6), rose and then fell in the second year (Time 7 to Time 13), then began another upwards incline in the third year (Time 14 to the end).

In fact, Theo was positive towards listening at the outset, telling me at Time 4 that it gave him '*a sense of pride*' and he was highly motivated to make progress:

'That's a random sentence that I've never heard before, yet I know it. What other ones could I know if I just keep going?'

Yet by Time 8, with Person Knowledge scores rising, he referred to longer passages as '*more intimidating*', and by Time 10, he was voicing a different perspective, saying:

'Listening is either way too easy then it gets boring, or it's really hard and then it's frustrating. I've never found an in-between.'

This reduced positivity – including a loss of his upbeat attitude – took place during the period discussed in section 7.6.2 which showed a strongly negative correlation between Theo's listening proficiency and his Person Knowledge scores. However, it might be

⁵⁹ Statements: 'I find that listening in French is more difficult than reading, speaking or writing in French' / 'I feel that listening comprehension in French is a challenge for me' / 'I feel nervous when I listen to French'

that a wider complex dynamic system had come into play: one which included Theo's view of his own learning and the role the teacher played in this. While at Time 5 he had told me *'I like it when I get things right,'* by Time 15 his perspective had pivoted from product to process, saying, *'now I'm listening so I can learn and improve, compared to listening to get a right answer. I'm taking in more things.'*

When asked at Time 19 to reflect on his experience of French learning throughout the study, he himself identified a change with a new teacher between Time 6 and Time 7:

'Ms K was a good teacher, but Mrs A did slower types of teaching and I didn't enjoy that as much because I wanted to race through and learn new things.'

Therefore, it might be that the return to growing Person Knowledge scores towards the end of the study reflected a new desire to improve born out of a deeper purpose to listening than simply getting the answer right.

7.7.2. A perfectionist learner relieved by the lack of control in listening

As a learner highly motivated to *'get the answer right'*, Theo initially identified a preference for reading and writing over listening because he could work at his own pace. However, even by Time 3 he had inverted this perspective. He identified that in reading and writing, he could manage his own time, and coupled with his ambitious nature, this brought out a perfectionism in him. The fleeting nature of listening, in which the speed and order of the micro-tasks are dictated by the speed and pace of the input, removed any potential to double and triple check, and this provided a level of relief. He told me,

'In writing tasks, you have to think about everything. You have to think about word order. When you're listening and just looking for what they're saying, it's a bit easier, I think, and I feel more relaxed.'

Despite relaxation, the fleeting nature of listening also caused excitement, summed up at Time 5:

'It's sometimes quite scary. Not scary, but it's a bit... you feel you have to feel focused, and you're on the edge of your seat and you've got your pen in your hand and you're ready and you have to be quick.'

The sense of calm and ensuing removal of control was again mentioned at Time 16 (over two years later), when, in comparison with speaking, Theo stated,

'when you're listening, you can be more passive. You don't have to think about pronunciation or anything, you just translate it.'

Furthermore, it appeared through our interviews that Theo very rapidly developed a strategy of moving on efficiently when he experienced difficulties with listening. This was discussed in section 7.4 from a linguistic perspective, but the complex dynamic system also took in elements of strategic behaviour. From the outset, at Time 2, he recognised that his first interpretation of a track might not always be the correct one, and talked about his inclination to hold both possibilities in his head while he came to his conclusion. And while he recognised the risk in this strategy at Time 3 to miss the next part of the input, he had learned by Time 9 to *'just move on'*. This was an approach he was still using at the end of the study, although even at Time 18 he recognised that it was easy to *'get hung up on a word,'* during the mental juggling act of making sense of the auditory input.

7.7.3. Translating word by word or chunk by chunk

I have discussed in earlier sections whether it was my paused translation methodology which inclined my participants to use English to make sense of what they understood

(e.g. Helen section 6.7.1 and Findings 1 section 4.3). I pushed Theo on this very point as early as Time 3, when we were discussing his paused translation, and he told me:

‘It goes French into my ear, then in my brain I’m thinking that means this in English, and then I put it back into French again... Cos I think it makes me think about it a bit more. When I put it back into English it’s kind of a check-point.’

In this early phase of the study, then, Mental Translation and Problem-Solving were working in consort for Theo to build meaning, as I illustrated in Figure 85, and his use of translation in suggests an inherent default towards the approach which was not linked to my methodology.

However, at Time 8, Theo told me that, *‘instead of thinking word-by-word I’m thinking more sentence-by-sentence’*. And given that the key statements in the Mental Translation element of the MALQ were ‘I translate in my head as I listen’ / ‘I translate *key words* as I listen’ / ‘I translate *word by word* as I listen’, it is logical that the relationship between the Mental Translation scores and those of Problem-Solving began to fall out of sync shortly after this time. In fact, Theo’s reports on his listening process were beginning to hint at an automaticity of listening, with translation still present in the repertoire of skills at his disposal, yet perhaps less foregrounded. For example at Time 9 he reiterated that, *‘I don’t think about the individual words, I think about it as a group,’* and by Time 10, *‘sometimes I’ll just keep it as French because that’s easier.’*

When Theo and I talked at Time 13, in July 2020, he had been remote-learning for four months due to the pandemic. The school had not offered live online classes; instead all learning was asynchronous homework-style tasks. He told me that he had done almost no

listening. His listening skills had been deteriorating and at this stage he had reverted to translating word-by-word.

However, on his return to school – with a return to far more exposure to aural French in timetabled face-to-face lessons, his automaticity of understanding rapidly came back: *‘I just hear it and know what it means.’* In the second lockdown at Time 17, Theo began to translate again, although this time:

‘Sometimes I’ll translate in blocks. If it’s slow, I’ll get one word at a time. But if it’s a block, I’ll remember the block in French and then I’ll translate it into English, but it’s not very automatic at the moment.’

Automaticity had returned to an extent at Time 20, depending on the difficulty of the input, and illustrates clearly how Mental Translation and Problem-Solving sometimes work in consort, and sometimes compete:

‘Sometimes it makes immediate sense, like I don’t have to translate the word et. Because I know that just means “and”. But if it’s a big word, or a word that’s important, then I’ll translate that, and I might have to think about it for a bit.

If it’s a hard sentence with lots of words I don’t know, I make sure to translate everything carefully, think about maybe substituting words. And then I have the sentence and I have blanks; I don’t know what the word means but there’s still more stuff coming in. I’m like, well there’s more stuff coming in that’s now more important.’

Theo’s own insights into his journey towards automaticity of understanding suggest a three-phase process, whereby the initial stage is a word-by-word translation into English, a second stage is a chunk-by-chunk translation into English, and the final stage is

automaticity of understanding without the need to translate. There was a similar yet reversed approach to Problem-Solving, given that as automaticity increases, there are fewer problems to solve. The two lockdowns in which Theo's exposure to aural French was vastly reduced aligned with a fluctuation between these phases, and his comments even at Time 20 indicated that the three phases faded into one another rather than being distinct. Theo's choice of stage would be dictated by the difficulty of the input.

7.7.4. Theo's metacognitive systems: concluding remarks

- Theo was markedly less anxious than the whole cohort throughout the study, with the exception of Times 10 and 11.
- His Directed Attention scores rose slightly throughout the study, while Mental Translation and Problem-Solving dropped slightly.
- The relationships identified in the full cohort whereby Problem-Solving and Mental Translation appeared to be a pair, as did Directed Attention and Person Knowledge, were not apparent in Theo's case.
- For much of the study there was a very strong correlation between listening proficiency and Directed Attention, although for the other metacognitive skills the relationships fluctuated more.
- Theo was a careful and strategic mental translator throughout the study who was mindful not to jump to conclusions in his initial understanding, preferring to test his hypotheses as his understanding built through a track. This was aided by his very quick thinking ability.

7.8. Theo's complex dynamics: conclusion

Although Theo, like Helen, developed highly proficient listening skills throughout the study, his trajectory is different. This difference is characterized by far more dramatic

fluctuations in listening proficiency as well as the linguistic skills of grammar and vocabulary, although like Helen's case, the ebbs and flows of grammatical and lexical knowledge appear to mirror each other over the three years.

Theo's approach to listening was efficient. He rarely allowed himself to become bogged down on vocabulary that he couldn't identify rapidly, preferring to move on and use other cues to build meaning. Rather than forming his own schema at the beginning of a passage, he would reserve judgement and skilfully use incoming information to develop or where necessary change his understanding of what he was hearing: a qualitative finding which is reinforced by the high moving correlation between his listening proficiency and his Directed Attention scores. Increased automaticity of understanding was apparent towards the end of the study but was highly reliant on increased input, as shown by the dips in automaticity during the lockdowns.

8. Findings 5: Fran

In exploring Fran's developmental trajectory of listening, I follow the same structure as I did with Helen and Theo. However, this chapter will demonstrate a very different developmental trajectory, informed by personality, motivation, and fluent Italian. As such, to an extent it challenges the model of listening strategies proposed in the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006): much of Fran's discussion of her approaches to listening does not map neatly onto any of the descriptors given.

8.1. Fran: Background information

Fran is an English girl living in the city with her parents and younger brother. She was recognised by those around her as being outgoing: she described herself to me at Time 5 as *'loud and bubbly'*, and reported that at parents' evening one teacher had said *'As soon as she comes into the room it's like a bomb has hit it.'*

As discussed in Theo's findings chapter, I had aimed for participants with no other languages and Fran's school data reported the language of the home to be English, but on interview she told me that she spoke mainly Italian with her parents, although mostly English with her brother. She also had some exposure to and understanding of Sicilian. She did some French at primary school, and had visited France two or three times in her life, although also would drive through France often on family holidays to Italy. Data acquired by the school on transfer to secondary placed her in the 'middle prior attainment' bracket. (Recall that at the participating school students in the 'low prior attainment' bracket were disappplied from language learning, so middle prior attainment is the less proficient of two groups.)

Of the four case studies, Fran had the most extra-curricular exposure to French. She spent several hours with a younger French girl while on holiday in Italy between Time 4 and Time 5, spent time with some francophone Swiss children at a summer camp in her home town between Time 6 and Time 7, and had a two-day stop-over in France shortly after Time 10 when driving to Italy again. In the February holiday before the covid lockdown she skied in France and attended French-speaking ski-school, which again included interaction in French and attempts at speaking and listening. In all of these interactions she acknowledged the communicative purposes of having an additional language, saying, *'to actually speak French to someone that's not in a French lesson, it's different. Cos you're not really learning French, you're producing it. You're putting it into action. And it's kind of nice.'*

Nonetheless, Fran frequently talked of her dislike of French lessons at school. At Time 3 she likened learning French to *'washing the dishes'* and at Time 5 identified the frustration of not being able to communicate as well in French as she could in English. This dislike seemed to have a range of reasons, and might have been tied up with challenging her self-image at the start of the study as a good linguist because she was bilingual. Sometimes it was because she was frustrated that the work was difficult for her when she compared it with her apparently effortless acquisition of Italian: *'People always say, "Oh Fran, stop moaning, you should find it easy," but I don't.'*

Fran began learning Spanish at the start of year 9 (Time 14), and found the decision of which language to continue to GCSE very difficult, although in the end she chose Spanish, and even continued this to A level.

In early March 2020, Fran was happy and confident, having returned from a family ski-trip and having just performed in the school play and won the school's 'Battle of the

Bands' competition. The covid lockdowns hit her sociable nature hard, exacerbated with particular concerns about her grandparents in northern Italy, and over the interviews that followed, she frequently spoke of her dislike of online school. She did not complete the Time 12 data collection round, saying that she had begun to find remote learning overwhelming; so much so that she was among a very small group of students who were invited back into school before the summer holidays for mental health reasons. With hindsight at Time 16 she recognised a loss of learning, likening her progress to a 'snakes and ladders' board:

'When lockdown hit, I think I was probably at about seventy, getting the hang of everything. Then it went down a snake to a fifty, maybe a forty. Then I kicked, and thought "oh god." Because we were being taught and I was thinking I know this stuff, why don't I get it? And then I really started understanding and I went up a ladder from a forty back up to a seventy and now I'm working up to a seventy-five, eighty. I didn't really refill in with what I forgot; I sort of learnt new things.'

Fran's thinking aloud at interview was characterised by rapid changes of subject, seemingly unconnected questions being asked of me, long monologues on her part, a need to move her hands and feet, and a collection of highly creative analogies to sum up her experience of listening. This behaviour was corroborated by reports of her experiences in the class and in the data collections, telling me at Time 3, *'I get distracted so easily, hands get sweaty, I look around the classroom, even the pitter patter of the rain.'* At Time 9 she spoke in depth about her need to fiddle, and use of strategies such as chewing gum and fiddle toys, and the importance of comfortable clothing. She had found the data collection that day particularly challenging, saying,

‘I was trying to fidget with something and there was nothing there. And I couldn’t go and pick something up and come back.’ I would be like, ‘wait, my brain is not working.’

Two connected themes came up repeatedly while I got to know Fran through our interviews. The first was a determination to excel, and a refusal to give up. This is best illustrated with her comment to me at Time 9,

‘I want to be the person who says, “no, I couldn’t have done better. I did all I could.”’

Linked to the desire to succeed was an extrinsic motivation born of others’ views of her: mine, her teachers’, her friends’ and above all, her family’s. All four grandparents were still alive at the start of the study, although two were very unwell, and her grandfather died around Time 16. Repeatedly she alluded to the desire to make them proud of her. At Time 5 (the week before her 12th birthday) she told me,

‘It’s really hard to keep going when you know you just want to sit there with them, and, like spend the last minutes with them. But I always think, if I fail – no, not fail, but if I don’t do well – what’s my Granny going to say. So I have to think about what she would say, and how proud she’d be of me.’

8.2. Fran’s listening proficiency and its variability

Fran’s final score on the listening proficiency scale was 62.85, which roughly equates to a CEFR score of A2. When correlating Fran’s listening proficiency scores with the time point, the correlation coefficient was 0.85, indicating a strong relationship between time and progress.

Table 49 and Figure 86 show Fran’s raw listening scores. Fran’s scores follow the mean very closely until Time 8, dipping at Time 9 but rising above average at Times 10 and 11 (on both occasions she reported being in France since the preceding test). Fran did not complete the test at Time 12, which was in May 2020 when the country was in full lockdown, but returned to participation at Time 13, where again her scores closely followed the mean until she pulled slightly ahead at Time 17.

Table 49: Fran’s listening scores relative to the mean

Time	Mean	Fran
1	36.40	37.03
2	40.06	45.24
3	40.26	45.31
4	44.14	46.40
5	41.85	47.34
6	41.59	41.23
7	47.42	49.79
8	41.99	44.64
9	44.62	39.92
10	47.45	55.07
11	48.87	54.40
12	49.07	
13	49.30	49.88
14	47.03	49.31
15	51.66	56.02
16	51.51	52.37
17	50.89	62.76
18	49.36	57.75
19	52.86	61.61
20	50.83	62.85

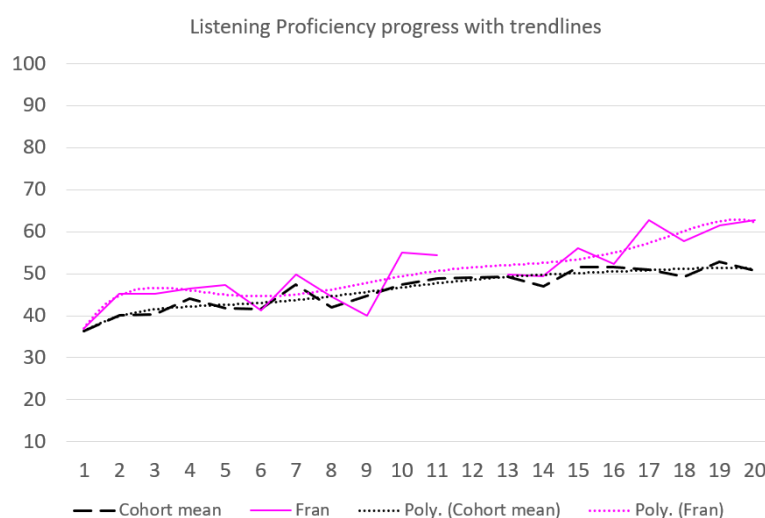


Figure 86: Fran’s listening scores relative to the mean

A moving min-max graph can be seen in Figure 87. Similar to Helen’s case, this graph illustrates steady progress in fairly small increments, yet with even smaller windows of variability than are seen with Helen, and marked differences with Theo’s fluctuations. Fran’s highest score was achieved at Time 20, although her Time 17 score was very nearly as high.

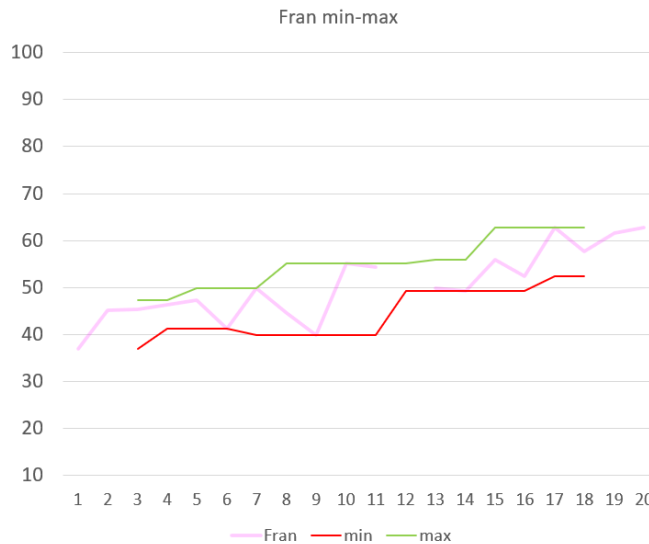


Figure 87: Fran's moving min-max listening scores

The change point analysis (Figure 88) software detected two change points in her learning: a first phase ended after Time 9 (from the start of year 7 to December of year 8), a second phase ran from Time 10 to Time 16 (from January of year 8 to December of year 9) showed progress despite the difficulties that came with the first covid lockdown, and the final phase from Time 17 until the end of the study. Each developmental jump detected by the software was around only 8 points on the scale, but with high confidence levels (first change 99% confidence level; second change 98%). (See Methodology section 3.6.3 for explanation.)

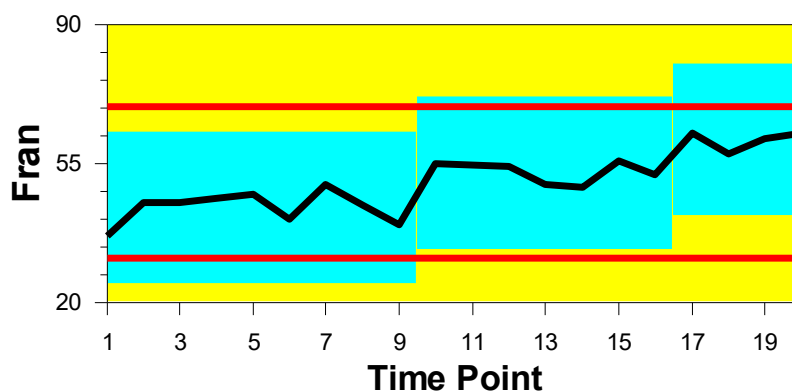


Figure 88: Fran's change points

8.3. Fran's linguistic subsystems, their development and variability

8.3.1. AuralLex scores

Fran's vocabulary scores as measured by the AuralLex test, compared with the mean, are given in Figure 89. Her vocabulary size measurements demonstrate much fluctuation: although the score began higher than the mean and in the main stayed there, she dipped at Time 8 and at Time 18, where she scored 0 (due to the large penalties for identifying non-words as words that she knew or could use). The four rotating tests are marked with the same colour in Figure 89 (i.e. the tests at Time 1, Time 5, Time 9, Time 13 and Time 17 were the same test, etc). Fran's largest vocabulary size of 1000 words was measured at Time 11 and at Time 17.

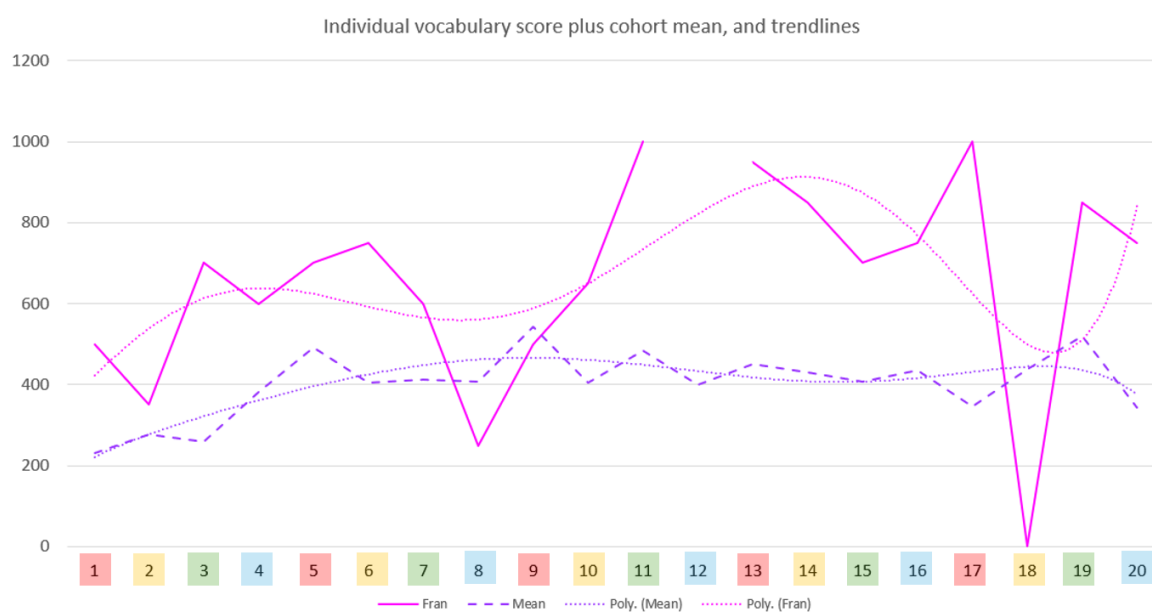


Figure 89: Fran's AuralLex scores, compared with the mean

Fran's vocabulary score as moving min-max is given in Figure 90. It illustrates the very wide range between minimum and maximum scores throughout the study, indicating an unstable system of vocabulary knowledge. The score of 0 at Time 18 – where points for correctly identified words were cancelled out by incorrectly identifying French non-

words – demonstrate that even towards the end of the study, Fran’s word knowledge was erratic.

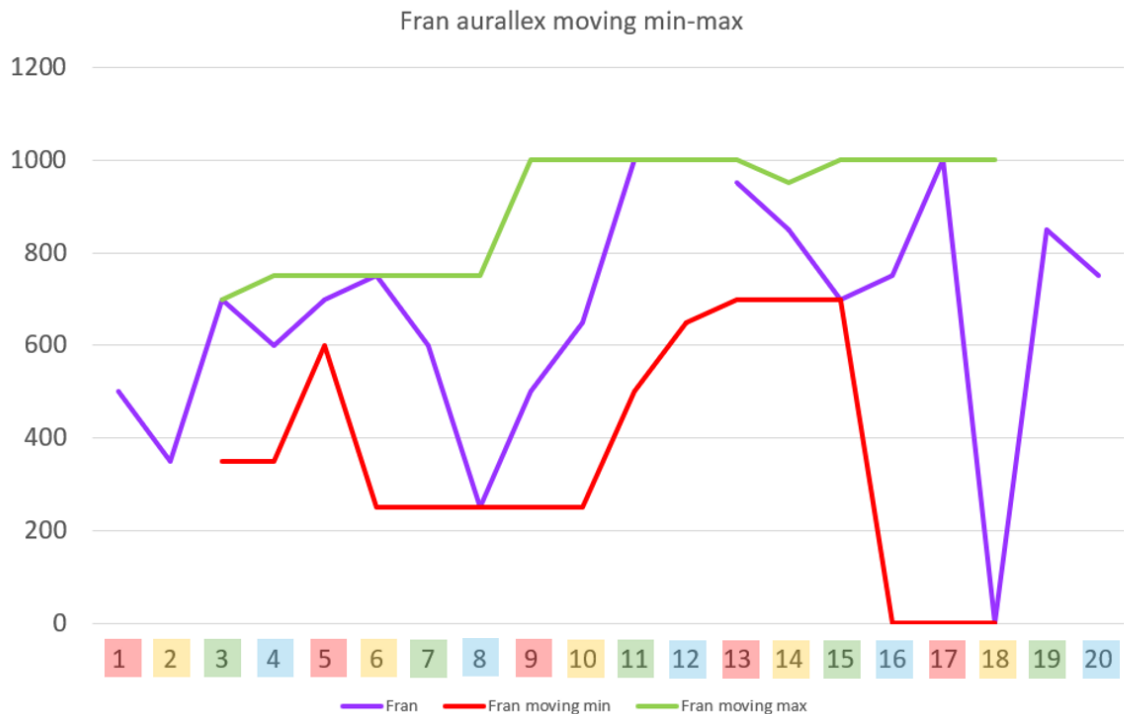


Figure 90: Fran’s AuralLex moving min-max score

8.3.2. AuraGram scores

The AuraGram scores for Fran, which measured grammatical development through an aural Grammaticality Judgement Test, are shown in Figure 91. (The repeated tests are colour coded) Fran’s trendline demonstrates a grammatical knowledge which began below average and which progressed at a rate in excess of the average, with her trendline crossing that of the cohort between Time 5 and Time 6, and remaining above average for the remainder of the study. Nonetheless, the individual scores show fluctuations which cannot be accounted for simply by the specific test.

Fran’s AuraGram moving min-max is illustrated in Figure 92 and further demonstrates the fluctuations and erratic performance in this measurement.

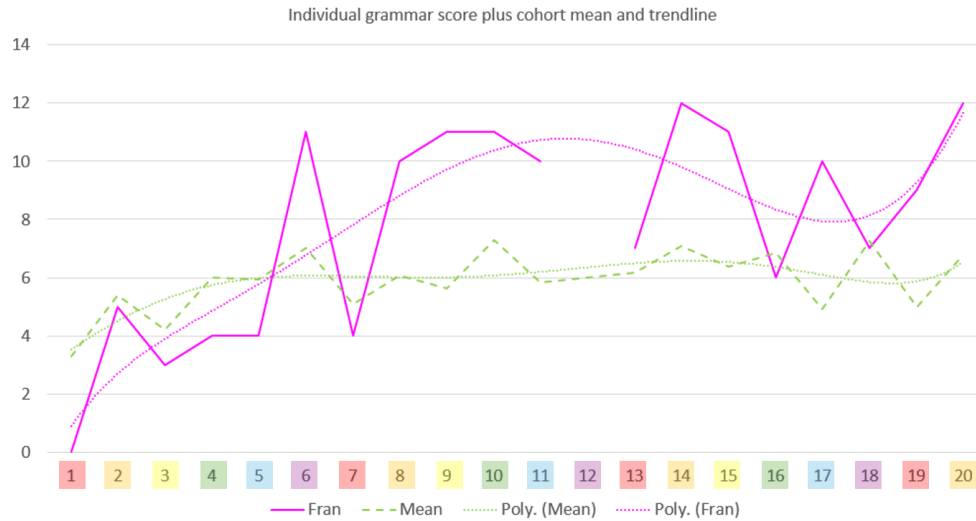


Figure 91: Fran's grammar scores plus cohort mean, and trendline

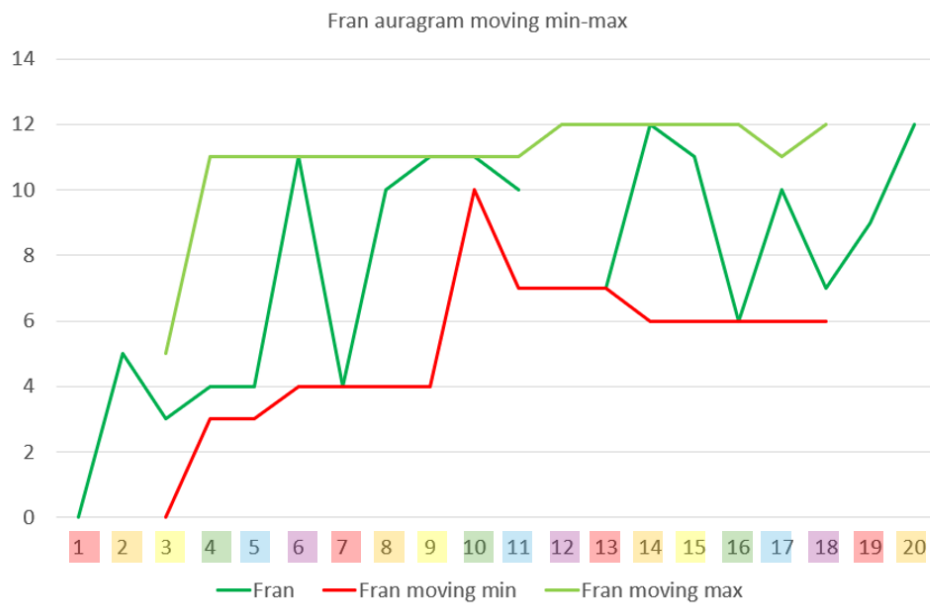


Figure 92: Fran's grammar scores moving min-max

8.3.3. Developing relationship between linguistic subsystems

Figure 93 places Fran's vocabulary, grammar and listening scores on the same graph, normalised on a scale of 0 to 1. Recall that the 0.5 line represents a vocabulary size of

600 words. In grammar terms, the maximum observed score in the cohort was 15/20, so the 0.45 line represents a score of 7/20. It is notable that in Fran’s case, her blue listening line is in the main lower than either of the linguistic subsystems, apart from grammar at the very beginning, and a slight dip in vocabulary knowledge (accounted for by the score of 0 at Time 18) towards the end of the study. This suggests that, as a rule, Fran demonstrated relatively good lexical and grammatical knowledge but lacked the ability to combine these skills and apply them together to decode the language she heard.

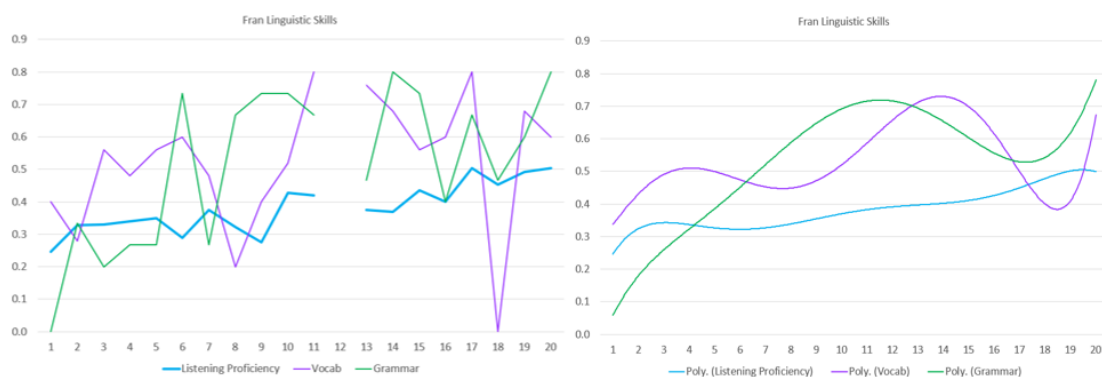


Figure 93: Fran’s listening, vocabulary and grammar scores plotted together

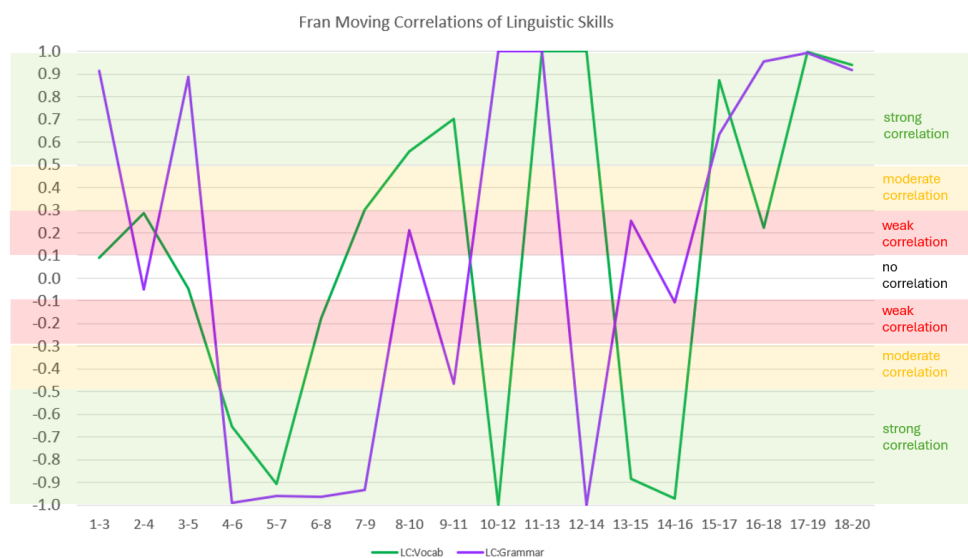


Figure 94: Moving correlations of linguistic skills for Fran

The moving correlations between listening proficiency and both vocabulary and grammar (Figure 94) indicate two highly variable relationships. It is difficult to give these relationships any clear categories (competitive, supportive, etc) given the rapid changes which occur from window to window, particularly in the central phase of the study – and it must be born in mind that the moving correlations are sometimes indicative of very small changes in the underlying systems. However it is worth noting that the only window in which there are consistently positive correlations are to be found in Fran’s final phase of listening proficiency as identified by the change point analysis software: in other words, from Time 17 onwards.

8.4. Fran’s errors and listening processes

Now I will explore Fran’s personal listening trajectory, in relation to the findings discussed above, where it is feasible to do so. For the whole cohort I found that errors might be driven by phonological or textual problems, although in reality there was constant overlap between these categories. Fran’s case is no different, although it leans heavily towards the phonological in terms of difficulties, and her understanding of French was both helped and hindered by her knowledge of Italian.

8.4.1. False friends that respect word boundaries

Similar-sounding words or phrases with different meanings – figured strongly in Findings 1. I noted in section 4.3.3 that the vast proportion of these were cross-linguistic (i.e. a French sound is heard as an English word such as *alors* / hello or *bouteille d’eau* / potato) rather than stemming from French vocabulary influence (*amuser* / *musée*).

Fran was very strongly inclined towards finding cross-linguistic pseudocognates throughout the study. There was no evidence of her erroneously detecting French words

within the speech stream and one case of a possible Franco-Italian pseudocognate which I will discuss within section 8.4.2, below.

Unlike the cohort at large, in the main Fran's pseudocognates respected word boundaries and usually made contextual sense. A summary of her errors is given in Table 50. It is important to note, however, that she was not always confident in what she had written in the paused translation tasks. At Time 18, when discussing *je me suis bien amusé* ('I had a lot of fun'), she demonstrated indecision or insecurity with her answer, switching from 'music' to 'museum' then back to 'music'.

Table 50: Fran's pseudocognate errors

Location	French	Correct translation	Fran's translation	Explanation	Word boundaries respected?	Contextual sense?
T6 E56	<i>Un port de pêche</i>	Fishing village	Fishing boat	'port' /pɔʁ/ mistaken for 'boat' (/b/ /p/ voice onset issue)	Yes	Yes
T7 F46	<i>Pays de Galles</i>	Wales	Portugal	/peɪdɔɡal/ sounds like 'Portugal'	Within phrase but not within word	Yes
T9 D25	<i>Tard</i>	Late	Tired	/ taʁ/	Yes	Yes
T10 D3	<i>Adore s'amuser</i>	Loves to have fun	Loves to go to the museum	~muser and musée (museum) are homophonous and musée is an Anglo-French cognate	No	Yes
T12 C17	<i>Pain grillé</i>	Toast	Pancake	/pɛ̃ ɡʁijɛ/ sounds like 'pancake' (/g/ /k/ voice onset issue)	Within phrase	Yes
T13 C55	<i>Avec lui</i>	With him	With Louis	'lyi/ sounds like 'Louis'	Yes	Yes
T13 C55	<i>Il est rigolo</i>	He is funny	Igloo	/ʁigolo/ sounds like 'igloo'	Ish	No
T14 D31	<i>Tu sais</i>	You know	You say	/sɛ/ sounds like 'say'	Yes	Yes
T16 C41	<i>Je déteste les choses</i>	I hate things	I hate shows	/ʃoz/ sounds like 'shows'	Yes	No
T16	<i>Bouteille d'eau</i>	Bottle of water	potato	/butej do/ sounds like 'potato'	Within phrase	Yes
T17 C11	<i>glace</i>	ice-cream	glass	/glas/ sounds like 'glass'	Yes	Yes
T17 C11	<i>Cher</i>	Expensive	Share	/ʃɛʁ/ sounds like 'share'	Yes	Yes
T18 St3AP	<i>On est allés</i>	We went	In Italy	/ɔ̃n ɛt ale/ sounds like both the French 'en Italie' and the English 'in Italy'	Within phrase but not within word	Yes
T18 St3AP	<i>Bien amusé</i>	Had fun	Good music	~musé and musique (music) are homophonous and musique is an Anglo-French cognate	no	No

8.4.2. Listening with Italian ears: a different set of challenges

Fran was unique among my case studies in being bilingual, and furthermore, being bilingual in a language very close to French. She initially perceived this as a disadvantage, telling me at Time 1, *'I keep getting things mixed up,'* and falling victim to some Franco-Italian false friends such as confusing French *gâteau* (/gato/, 'cake') with Italian *gatto* (/gattɔ/, 'cat') and French *beaucoup* (/boku/, 'a lot') with Italian *bocca* (/bɔkka/, 'mouth'); the latter word causing problems for much of the study.

Although I have placed this section within the linguistic subsystems, there was also an affective element to the role Italian played in her making sense of aural French input. At Time 2 she referred to *'just Italian words flashing through my head,'* and talked about it *'blocking everything.'* At this early point in her learning of French she felt as if she had to seize moments where the 'block' was lifted and work hard before it returned:

'There are times when I'm like look, it's gone, Fran! Come on! Come on quick!'

Yet at the same time point, in her stimulated recall she would give the translation from French first into Italian, and then into English. By Time 3 she told me she was *'learning to block out the Italian,'* but also recognised its advantages, recalling a classroom incident where she was the only student to be able to translate *faire* ('to do'), given its proximity to the Italian equivalent *fare*, even noting her superiority over case study Theo in this regard.

This ebb and flow of the Italian both helping and hindering continued throughout the first half of the study: the analogy of the 'block' developed into a 'wall' at Time 5, which she could sometimes see through, but at other times not, and couldn't predict when. At Time

10, it had become a ‘door’. Although she said her default was to translate into English, it appeared that there were occasions in which French-Italian cognates were of help:

When they said monde, I was like “oh okay, mondo”, was the first thing that popped into my head. “Okay, that’s world.” It’s like a door. The Italian would be like “hellooo, this is what it means!”, and then it’ll be like “bye-bye.”

A similar approach was seen at Time 5 with *ma passion* which became *mio passione* [sic] and Time 11 where *jardin* was *giardino* before it was translated into ‘garden’, and *notre* was *nostra* before it became ‘our’.

The change point analysis pointed to a second phase of listening proficiency between Time 10 and Time 16 for Fran. During this phase, Fran demonstrated progress with *beaucoup* / *bocca*. Our stimulated recall at Time 13 exemplifies this:

Il a beaucoup d’énergie. And you wrote ‘he has a lot of energy.’

That’s surprising. Mouth. Beaucoup means mouth.

You heard Il a beaucoup d’énergie and you’ve written, ‘he has a lot of energy,’

You didn’t write anything about mouths.

Oh no, sorry. I was thinking Italian. Mouth is /bɔ̃kkɑ/. It’s a lot. I knew it as soon as I did it. I thought, “oh, we’re talking about French, we’re not talking about Italian.” I needed to switch.

Still, there were numerous occasions during our stimulated recall where Fran could not use her Italian to access French. At Time 12 she was unable to translate *lait* (‘milk’ / *latte*), *mettre* (‘to put’ / *mettere*), or *miel* (‘honey’ / *miele*), and at Time 16 the same was true of *vrai* (‘true’ / *vero*). This indicates that the relationship between fluent Italian and easy access to French is not straightforward within the stream of speech.

In Findings 1 I demonstrated similar examples for the whole cohort between French and English (such as ‘spa’ from *ne m’intéresse pas*). At Time 18, Fran also gave one single example in the dataset of a complex French-Italian pseudocognate, which crossed word boundaries and interfered with her understanding. Her translation of *l’année dernière* /lanedɛʁniɛʁ/ (‘last year’) was Monday, which could be explained when one considers that the Italian word for Monday is *lunedì* /lunedi/. It also points to the subtlety of vowels and the salience of consonants in aural comprehension, as addressed in Findings 1 section 4.3.3.

From Time 15 onwards Fran had discussed with me how easy she was finding Spanish, thanks to her Italian knowledge, and she was no longer negative about the Italian influence. By the end of the study, Fran told me:

Sometimes I repeat the words in my head to try and think of cognates or something in Spanish or Italian or English... if you were to take away those from anyone, I don’t think they’d be able to do it as well.

She had recognised that her knowledge of other Romance languages was an advantage.

8.4.3. Plurals and tenses: morphological features are not detected

In Section 4.5.1 of Findings 1 I discussed the differences between content and function words, and I also discussed Theo’s case in 7.4.3. In Table 51 I reproduce Table 38 in percentages to enable comparison with Fran’s experience: Table 51 demonstrates that Fran’s total difficulties (see column C) are broadly in line with the cohort’s averages. However, Fran was much more likely to omit a content word than to submit an incorrect translation. This is noteworthy given that her AuralLex scores were consistently above average; it is consistent, then, with the finding that her overall listening proficiency was

relatively lower than her lexical proficiency. In other words, for Fran, factors other than vocabulary playing a role in impeding her comprehension.

Table 51: Missing and wrong function and content words: Comparing Fran with the full cohort

	A Missing		B Wrong		C Total	
	Fran	Cohort	Fran	Cohort	Fran	Cohort
Content word	45.3	20.1	34.9	57.9	80.2	78.1
Function word	17.4	19.3	2.3	2.6	19.8	21.9
Total	62.8	39.5	37.2	60.5	100.0	100.0

Function words also posed a problem for Fran, in line with the average. Therefore it makes sense that morphological features also posed a problem, and this became clear analysis of Fran’s errors. (Furthermore, we have already seen that these were problematic for the more proficient listeners Helen and Theo.) Throughout the dataset Fran failed to recognise plural markers: for example at Time 8 erroneously expecting to hear a final /z/ sound to indicate plurality as in English:

‘It’s only one brother, because it says /avɛk me frɛ/ but it would be /frɛz/ if there were two.’

By contrast, at Time 14, *ma cousine* was translated as ‘my cousins’.

A similar picture can be drawn with Fran’s failure to recognise tense markers in verbs: in fact correct tense did not appear to occur to Fran when she was justifying her translations, preferring to use context. I hypothesised in Findings 1 section 4.4.3 that such a perception might be rooted in a transfer of training, and when pushed at Time 14 as to why she had translated a reference to holidays in the past tense, she justified this by saying: *‘I think it was because after a holiday you come back and they ask what did you do. I didn’t think about it.’*

8.4.4. Common sense fights with perception

I showed in Findings 1 that ‘sound trumps sense’: in other words, that participants favoured translations that reflected the sounds they heard, regardless of logic.

Fran’s answers often lacked sense, but her comments indicated that she was grappling with this issue. At Time 5 in response to *Les chips? J’adore les chips*⁶⁰ (see Figure 37), she told me:

*‘It was trying to say “chips,” but it sounded like they were saying “sheeps”?
So I was writing sheeps! I love the sheeps! It was like “I love to eat sheeps,”
so I was really confused. Because what does it mean by you eat sheeps?’*

Her comment to me regarding her translation of *rigolo*⁶¹ /ʁigolo/ as ‘igloo’ might shed light why ‘sound trumps sense’. Although she, like five others, chose this nonsensical translation, Fran said, *‘I know it’s not “igloo”, but it sounded like it.’* This raises the possibility that other participants also knew that what they had written was incorrect, but wanted to record their impression of what they had heard.

8.5. Summary of findings for the linguistic systems

In terms of Fran’s developmental trajectory, then, we have seen:

- A small but significant increase in listening proficiency, which follows the cohort average until Time 17. Change points were detected before Time 10 and Time 17.
- Very erratic vocabulary and grammar scores, which were nonetheless on average higher than the cohort average.

⁶⁰ Track 4 test F28

⁶¹ tracks 2 and 4, test C55, T13

- Even despite the erratic fluctuations, Fran appeared to be better at grammar and vocabulary individually: she has the individual proficiencies but appears to lack the ability to exploit these skills for use in aural decoding, and there was no evidence of this grammatical ability within the paused translation dataset.
- A better than average perception of word boundaries, but nonetheless a struggle with Anglo-French pseudocognates.
- Some recourse to Italian in her translations, particularly of Franco-Italian cognates, which she initially resisted, but eventually recognised as advantageous.

8.6. Fran's metacognitive subsystems, their development and variability

8.6.1. Metacognitive scores

Fran's scores for the metacognitive subsystems are given in Figure 95, relative to the mean, with moving min-max scores in Figure 96. Recall that the statements were rated on a six-point Likert scale, where 'strongly agree' would equate to a score of 1.0. The MALQ is in appendix 15.1.1.

Figure 95 shows a drop in Person Knowledge⁶² from Time 3 to Time 11, after which time the trendline is fairly stable. By contrast, Fran's Directed Attention⁶³ followed a reverse pattern, whereby after Time 5, she generally 'agreed quite a lot' to these statements. The dip she showed at the end of the study towards 'agree a bit' aligns with the second listening proficiency Change Point which was found to be at Time 17.

⁶² listening in French is more difficult than the other skills / listening comprehension is a challenge / I feel nervous when listening

⁶³ I focus harder when I have trouble / I recover lost concentration rapidly / I try to get back on track / I don't give up when I have difficulties

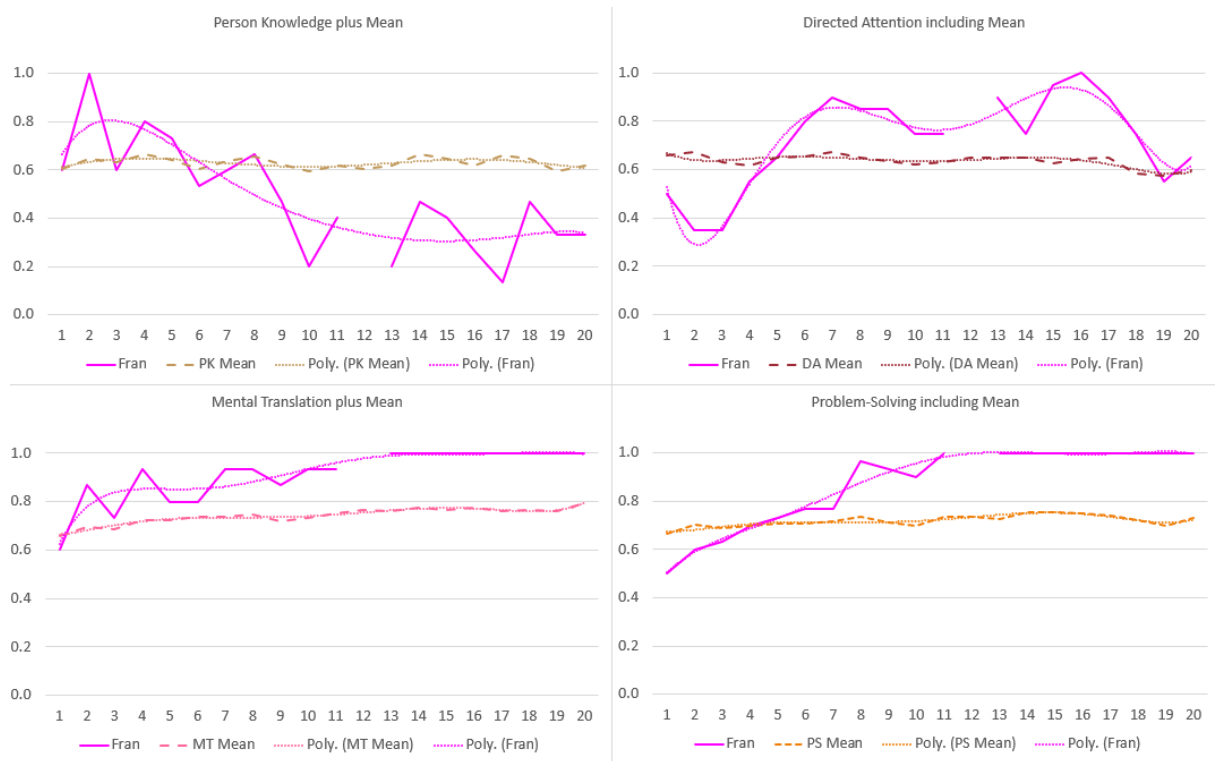


Figure 95: Fran's individual metacognitive scores relative to the mean

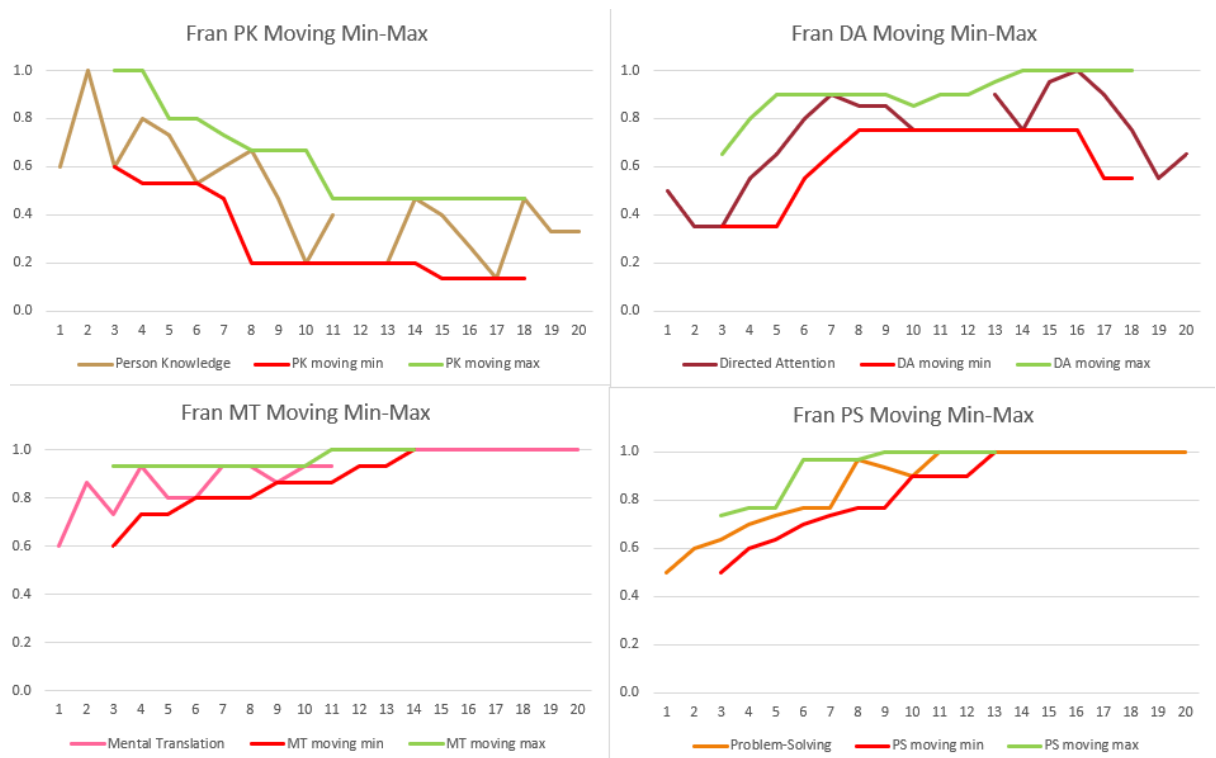


Figure 96: Fran's metacognitive moving min-max scores

What is perhaps most striking, though, is the consistency of responses to the Mental Translation⁶⁴ and Problem-Solving⁶⁵ statements after Time 12: although the statements were randomised within the survey, she fully agreed with all the statements relating to these constructs, claiming that she always translated in her head and used context thoroughly to check her understanding. In section 8.7, I will probe these assertions by triangulating them against the qualitative data provided through the stimulated recalls and paused translation.

The moving min-max graphs in Figure 96 reinforce the finding above that Fran's listening anxiety began to dip from Time 3, but highlights the large drop in minimum score from Time 8, followed by a large drop from Time 10 in maximum score, Time 10 being the first change point found in Listening Proficiency. Like in Theo's case, we can see a narrowing of the Directed Attention moving min-max bandwidth, but unlike Theo, this widened again from Time 16, pointing to renewed instability in this system. As discussed in the previous paragraph, the bandwidth for Mental Translation narrowed to a point at Time 13, and for Problem-Solving at Time 14: for Fran (according to self-report at least), these two subsystems became entirely stable and reliable until the end of the study.

8.6.2. Developing relationships between metacognitive subsystems

In section 5.3.2 of Findings 2, I showed that on average, trajectories of Mental Translation (pink) and Problem-Solving (orange) followed each other, and that the same was true of Directed Attention (burgundy) and Person Knowledge (beige). Figure 97,

⁶⁴ I translate in my head as I listen / I translate key words, I translate word by word

⁶⁵ I use the words I understand to guess the words I don't / I compare what I understand with what I know about the topic / I use my experience and knowledge to help me understand / I adjust my interpretation if I know it's not correct / I use the general idea of the text to help me guess / I check back to see if my guess makes sense

which gives all Fran’s metacognitive subsystem scores on a single graph, shows that Mental Translation briefly aligned with Person Knowledge, but then these two separated and instead Mental Translation fell into step with Problem-Solving.

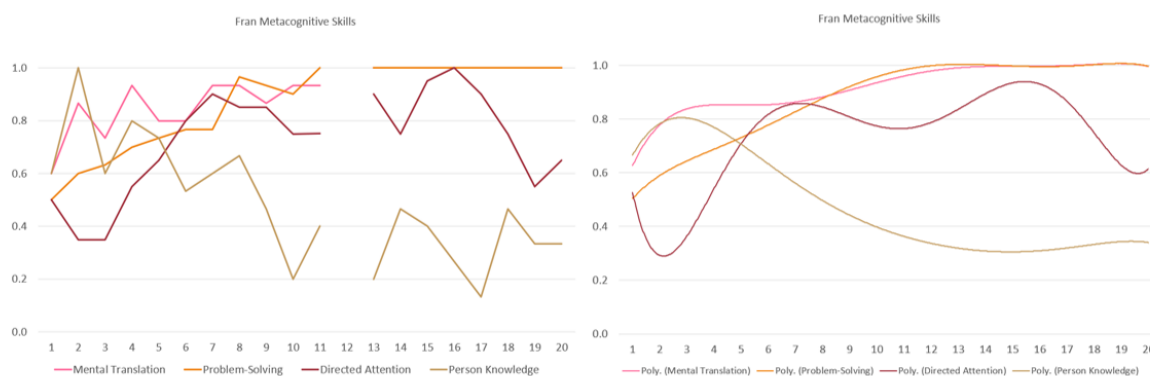


Figure 97: Fran’s metacognitive scores on a single plot: raw scores and trendlines

Figure 98 illustrates the moving correlations between Fran’s listening proficiency and each of the four metacognitive subsystems. With the exception of the Time 2-4 window, the moving correlation between Person Knowledge and listening proficiency remained positive – and mostly strongly so – until the Time 7-9 window (this matches the first phase of listening proficiency according to the Change Point Analyser). Given that the Person Knowledge subsystem aims to measure self-reported listening anxiety, this suggests that initially, there was a strong positive correlation between how anxious she felt and how well she performed in the listening proficiency tests. There then followed a brief period where the reverse was true, ending at the Time 10-12 window, and a period of fluctuation. However, from Time 15-17 until the end of the study, there was a mostly strong negative correlation between the two variables; in other words, by this time, Fran achieved higher listening scores when she reported being less anxious.

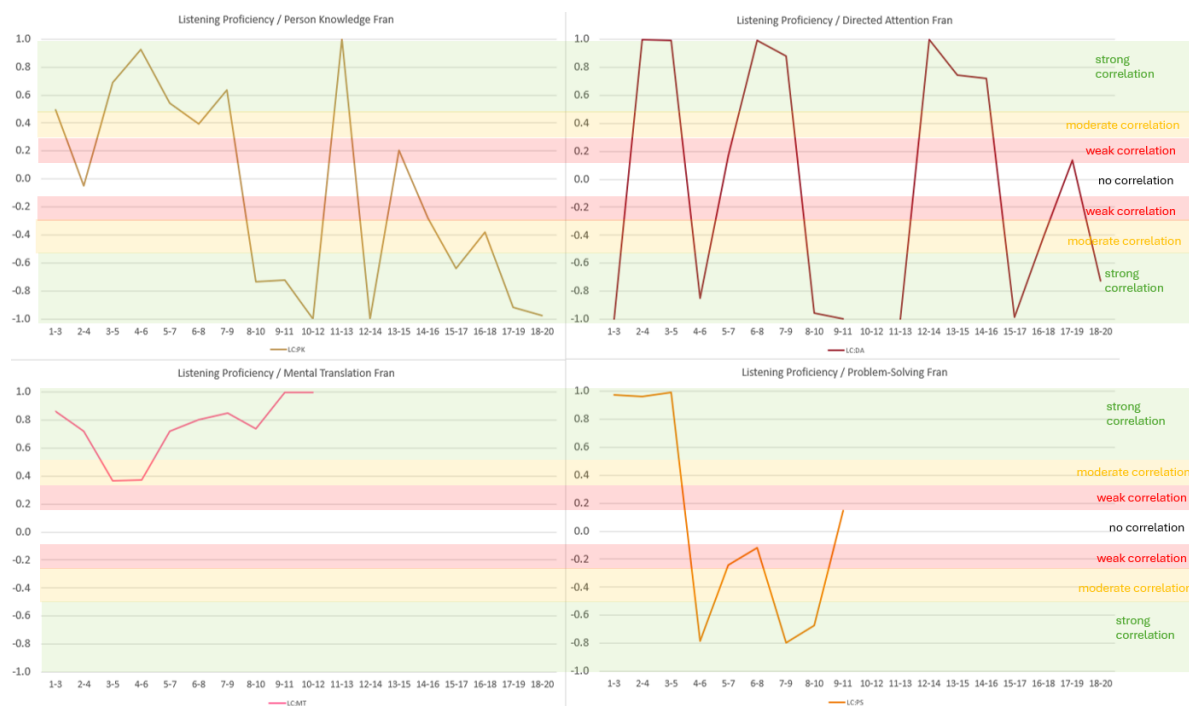


Figure 98: Moving correlations between listening proficiency and the four metacognitive subsystems

The repeated fluctuations in the moving correlations between listening proficiency and Directed Attention mean it is very difficult to draw any conclusions, and moving correlations cannot be calculated at all for the final time points of Mental Translation and Problem-Solving due to the consistency of scores (since the correlation calculation relies on finding a standard deviation between individual measures). However, in the initial phases of the study there was a strong positive correlation between listening proficiency and Mental Translation, and wide fluctuation between listening proficiency and Problem-Solving.

8.7. Fran's strategic journey

Findings chapter 2, and the second parts of Theo's and Helen's case studies discuss participants' metacognitive developmental trajectory largely through the lens of the categories designated by the Metacognitive Awareness Listening Questionnaire. Fran's own developmental trajectory, as elicited at interview, does not map so neatly onto the

MALQ; our conversations during interview often inclined more towards self-regulatory strategies, self-efficacy, and motivation. In order to address the question of relative weights of different subsystems at different time points, the four subtitles which follow superimposed on Fran’s listening trajectory in Figure 99.

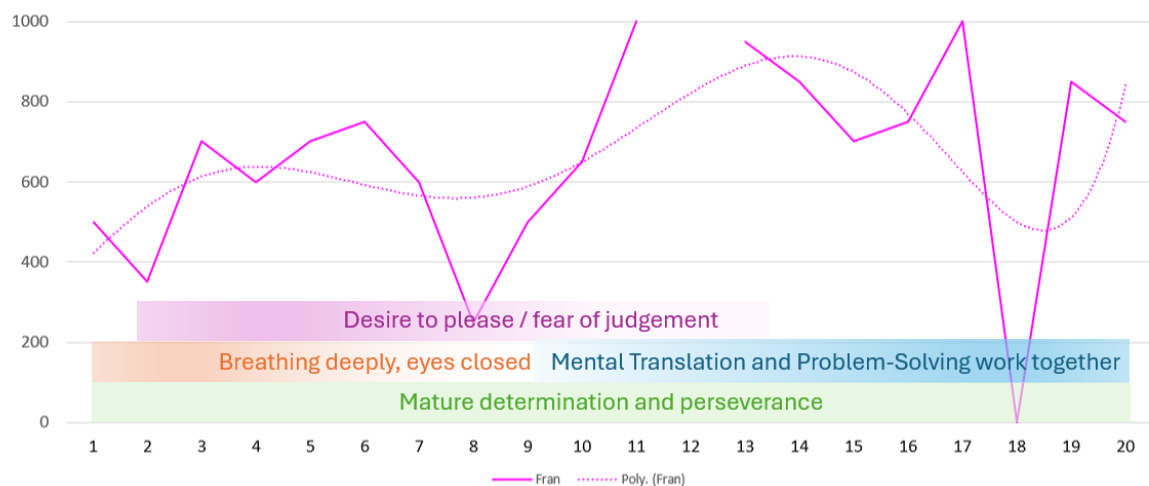


Figure 99: Fran’s affective approaches over time

8.7.1. Breathing deeply, eyes closed to keep calm

When I asked Fran about strategies used to help her listen, her answers between Times 2 and Time 9 revolved around themes of self-regulation. She frequently referred to removing her headphones during the data collection and taking ten seconds to focus on her breathing before returning to work. Although she rarely alluded to a lack of confidence or specifically to anxiety surrounding listening, it was clear that this was part of an inventory to boost confidence; at Time 4 she stated:

‘I close my eyes for ten seconds, then open them and have a look at the keyboard and, every time I focus on a different letter. I look at that letter, and I think of a word in French that starts with that letter. And when I do that, I’m just like, “OK, so you’ve thought of that on the spot! OK, so you’re actually not bad at this.”’

Such a strategy suggests that she recognised a discomfort with the study's listening tasks, and had recourse to address them through calming strategies. And despite the general dislike of French discussed in section 8.1, Fran began to talk to me positively from Time 6 about in-class listening tasks, which often involved matching pictures with short audio tracks. She reported enjoying the '*funny voices*,' but at Time 8 also referred to the inevitable quiet that would come over the classroom preceding the beginning of a task. At this time Fran also referred to a series of self-soothing strategies that she would call upon to ensure maximum relaxation:

'If you just breathe, don't do anything, put your feet on the ground, and just listen. And then... you're more relaxed, it's easier to think than if you're panicking.'

This developed by Time 9 into a feeling of pleasure initiated by the audio announcement of a listening task, which would begin with a French announcement of the page and exercise number; an excitement to which she frequently alluded in subsequent interviews.

8.7.2. A desire to please and a fear of judgement

Regardless of the enjoyment of in-class listening exercises, Fran also told me she disliked my paused translation tasks, initially born out of a misunderstanding of her role in the process. She told me that she had said to herself at Time 2,

'It's for research, you'd better get this right cos people are going to judge you... then I felt quite upset because I put "I don't know" for literally all of them.'

Furthermore at Time 3 she expressed worry that she was not going to be helpful because she was '*doing terribly*,' and at Time 4 alluded to the need to please her teacher and nagging doubts about her own insufficiency.

‘But when I panic, it’s like OK Fran, you need to well, because Miss could give you an F, and I know that that’s nothing to do with this, but that’s the instinct that kicks in. That – not just that I’m going to be judged, but Miss is going to – I’m a failure.’

These anxieties extended beyond French and her participation in the research: she discussed the desire to excel in other subjects, and once the pandemic lockdowns began, experienced anxiety about the implications of not getting all the work done. Ironically shortly before Time 12, these anxieties became paralysing, and very little lockdown work was done.

8.7.3. A mature and big-hearted attitude to learning

Recall that Fran, like all the participants, had recently turned 11 when the study began, and turned 14 in the final academic year of the study. Despite her anxieties (or perhaps because of them) she showed a remarkably mature and sanguine approach to learning which will largely fall into the MALQ’s Directed Attention⁶⁶ subsystem, but also taps into wider pedagogic theories of motivation and self-efficacy. Her underlying response to the anxieties verbalised in section 8.7.2 was *‘I will be able to do it in the end,’* (Time 1); *‘Dedication and perseverance; I will get better at it,’* (Time 3), carrying on throughout the study to *‘It’s nice to me to think at least I’ve tried,’* (Time 16) and *‘what’s the point in me not getting back in and trying? What does that show? It shows that I’ve just given up,’* (Time 20).

Fran worried that I would judge her and said that although the test battery was stressful, but she very much enjoyed the interviews. She was also highly motivated by

⁶⁶ I focus harder when I have trouble / I recover lost concentration rapidly / I try to get back on track / I don’t give up when I have difficulties

participation as she mused upon the potential impacts of the present study and her role in changing French pedagogy, showing a maturity of insight and altruism.

8.7.4. 'You press a button in your head': Mental Translation combined with Problem-Solving

In initial interviews, as discussed above, talk of strategy tended to default to self-regulation and breathing techniques. Time 9 was the first in which she explicitly mentioned translation, which then became a recurring theme. In section 8.6.1 I showed that from Time 13 onwards, Fran 'agreed completely' to all the Mental Translation questions⁶⁷ from the Metacognitive Awareness Listening Questionnaire, and from Time 14 to the Problem-Solving questions⁶⁸. I had wondered whether this had been due to test fatigue, but concluded that it was not: firstly, the questions in the MALQ were randomised and Fran did not 'agree completely' to every single statement of the MALQ, and secondly her comments at interview are consistent to demonstrate a commitment to translation in order to build understanding.

The Mental Translation and Problem-Solving subsystems appear to offer the best insight into Fran's mental processes while listening, particularly in the second and third phases of her trajectory (as detected by the change point analysis). At Time 9 she alluded to using her echoic memory as well as problem-solving strategies when she told me,

⁶⁷ I translate in my head as I listen / I translate key words as I listen / I translate word by word as I listen

⁶⁸ I use the words I understand to guess the meanings of the words I don't / As I listen, I compare what I understand with what I know about the topic / I use my experience to help me understand / I adjust my interpretation if I realise it is not correct / I use the general idea of the text to help me guess the meaning of words I don't know / When I guess, I think back to see if my guess makes sense

‘I’ll try and think of the word in my brain, but I’m not going to write anything, I’ll just think it, and I can still hear the other bits. Sometimes I’ll wait until the whole sentence is up, and then I’m like, “what can that mean?”

Such an approach was well used when she was asked to translate *A Paris on peut faire du tourisme* (‘In Paris one can do tourism’). She said, ‘In Paris there is Disneyland’ and explained that she had focussed on the word *faire*, and, consistent with the findings cited in section 8.4.1, seized on it as a false friend ‘fair’, then applying her world-knowledge, thought of the one fairground she knew about in Paris: Disneyland. (This is despite the fact that she had correctly – and proudly – translated ‘faire’ at Time 2).

A similar combination of context and use of cognates (a correct Franco-Italian cognate this time) led her at Time 10 to translate *tout le monde est formidable* (‘everyone is great’) as ‘to the moon and back’. It had followed the sentence *J’adore ma famille* (‘I love my family’), and she told me,

‘She said something about monde, and mondo in Italian is the world. “The earth and back?” But we’re on earth... Well, mondo in Italian means world, and most mums, they love their kids to the moon and back, obviously, but I don’t know if French people say that.’

Despite these explanations of how she had given meaning to what she had heard, Fran struggled to put her experience of listening into words (‘I have no idea what’s going on,’ (Time 14); ‘It’s one of those questions like, “how do you breathe?” I have no idea how you want me to answer that question,’ (Time 17)). At Time 15 she embarked on a series of analogies to summon up the experience of semi-comprehension:

'There's a start, a line, and an end. So the word, or the sentence, can either travel across the line and get to the end and I understand it, or it could start the line, wobble, and fall off. If I understand it, it might go and wobble and keep going. Or it'll fall off and it'll fall back on and keep going because I can fill in the gaps.'

And by Time 17, the idea of 'falling off' had been replaced by a concept of rearrangement of words, fuelled by a determination to succeed reminiscent of Directed Attention, as well as Problem-Solving strategies:

'I feel like I concentrate harder when I don't understand it. It's a bit like magnets on a fridge, like an anagram. You put them out and then you rearrange them and then you rearrange it until it sort of makes sense but there are gaps. Then you fill in the gaps with your common sense.'

The final analogy, at Time 20, was that of a sentence as a washing line:

'And there are loads of socks on that washing line. And every two socks, there's four that have fallen off. And I write down the words that are where the fallen socks are. And I try to figure out how I can put those socks back on the washing line by trying to think of words that can go in there to make sense.'

Throughout all the rearrangements, the filling the gaps, and the wobbles, though, there was an underpinning approach of understanding by means of explicit translation. Fran reported progress in that she felt her translations were more accurate, but they were, nonetheless, translations.

8.7.5. Fran's metacognitive subsystems: concluding remarks

- Listening to French was a strongly affective experience for Fran, which touched on concepts of self-regulation, self-efficacy and motivation as well as metacognition.
- There was a brief initial alignment of Mental Translation and Person Knowledge, whereby there was both more anxiety and more translation, and this was confirmed by much initial allusion to anxiety during data collection.
- Mental Translation and Problem-Solving aligned from Time 8. Here there is evidence that Fran uses both subsystems together to form meaning from the input.
- There was a strong and positive moving correlation between Mental Translation and listening proficiency.

8.8. Fran's complex dynamics: Conclusion

Fran's developmental trajectory of listening illustrates a third path in contrast to those of Theo and Helen. Fluent Italian may have boosted individual grammar and vocabulary scores yet without contributing directly to overall listening proficiency. Added to this was the profoundly negative impact of the covid lockdowns.

Although Fran's listening proficiency scores closely resembled those of the mean, her approach to making sense of what she hears were founded on a unique combination of a determination to succeed fuelled by anxiety and fear of judgement but also on a personal inventory of self-regulatory strategies to allay these fears, and ultimately make progress.

9. Findings 6: Ryan

In this final case study, I explore Ryan's experience. It follows the same structure as the previous three case studies, but demonstrates another trajectory quite different both from the full cohort, and from Helen, Theo or Fran.

9.1. Ryan: Background information

Ryan (a pseudonym) is a monolingual English boy who was adopted by his grandparents as a toddler. The three of them live a sociable and outgoing life in a village just outside the city where Ryan enjoys roaming with his friends and making dens. He is also a keen footballer and figure-skater, and played the cornet.

Ryan had done a small amount of French at primary school, but had no exposure to French outside the classroom, and did not visit a French-speaking country during the data collection period. Ryan did not choose to continue his study of French after the statutory language-learning period was over at the end of year 9 (Time 20).

Data acquired by the school on transfer to secondary placed him in the 'middle prior attainment' bracket. Although Ryan was chosen as a middle-attaining boy at random, he was among the least proficient learners in the sample and his listening proficiency scores were in the bottom ten per cent. During interview he was sometimes incoherent; not only did he lack vocabulary in his native English, but he seemed sometimes to omit key points in his narrative, so that what he was saying might have made sense to him but did not to me. The following exchange from Time 18 illustrates this well:

So, how are you feeling about the idea that you're not going to be doing French any more after year 9?

What's the word?

Happy? Sad? Relieved? Excited? Disappointed?

A bit of everything.

Because?

Because I'm not really thinking about going on a holiday to France. But it's still useful at times, if I were to go to a place that wasn't France, but people still speak French.

When asked about his learning in French lessons, Ryan was also often unclear. At Time 10 I had expected a comment about topics being covered, but instead he told me, (probably in reference to a gap-fill exercise) *'we're filling in spaces, I don't know, I was so confused.'* He regularly told me that he had learned nothing new since year 7 (i.e. Time 6), and that all subsequent lessons were concerned with revision, and yet, at Time 18 that the revision *'hasn't gone in because it's been different.'* When asked at the end of the study about what he felt he had learned in French he said,

'I can say I like... foods. Tiny bit of, like, countries. Numbers. Not alphabet because it's... I just can't get that right. Um, ... I can say how you're doing, and I can reply to that.'

While this was probably an accurate representation of his productive skills at the end of his French education, he also told me that he was good at listening in class; the accuracy and nuance of this will be explored further in section 9.7.2.

Ryan seemed to take the covid closures in his stride, but he was disorganised with school work. In fact, he failed to complete so much of it in the initial phase of lockdown that he was invited to come into school – only ten students (in a school of over 1,000) had

completed less than he. Inadequate work became something of a theme; on return to school at Time 14 he began to collect a number of detentions for incomplete classwork or homework, and during the second lockdown from January to March 2021, Ryan was asked to attend school nearly every day (other than a day's exclusion for bad behaviour).

Despite Ryan's outgoing nature, he sometimes encountered social problems at school; he struggled in the change of groupings between year 7 and year 8 (Time 6 to Time 7), telling me he had no friends in the new class, and the subsequent change between years 8 and 9 (Time 13 to 14) was so unsuccessful due to bullying that he changed classes again at Time 17. A second lockdown coincided with this change of classes, and he attended school for supervised remote learning (lessons still taking place via Google Meet), telling me he was very happy to be there. An upshot of the change, though, was that despite being in a classroom with a teacher, he still missed a number of online lessons, claiming he did not know his new timetable, but also told me that he had '*done most of the work.*'

Ryan's own accounts of events sometimes contradicted other evidence and contradiction was a characteristic of many of our interviews. It was clear that Ryan lacked critical thinking skills and at around Time 15, he was diagnosed with a special educational need, which consisted of weak working memory skills, forgetting instructions, frequent loss of focus, difficulties with reading comprehension and a tendency to claim that he understood tasks when this was not the case. All of these difficulties are likely to have influenced what he said to me, and in the present analysis I present both Ryan's words, and, where appropriate, data collected through other means, in an attempt to explore Ryan's developmental trajectory. Sometimes it appeared to me that he was simply feigning understanding, and at other times I wondered whether, referring back to his special education needs diagnosis, his 'tendency to claim that he understood tasks when this was not the case' was even a claim to himself – in other words, a tendency towards

self-deception. For example telling me after a paused translation task where he had managed to translate around seven words from a 90-word track that he had '*found that really fine.*'

Given the unreliability of his narrative, I gave serious consideration to whether or not to present Ryan's case in this thesis. However, pupils such as Ryan are not uncommon, yet their experiences are underexplored in academic literature. The analysis of Ryan's developmental trajectory and his unique complex dynamic system brings with it a wealth of opportunity to explore both pedagogical and methodological implications and these will be discussed in chapters 10 and 11.

9.2. Ryan's listening proficiency and its variability

9.2.1. Variability in listening proficiency scores:

Table 52 and Figure 100 give Ryan's raw listening scores. They show a proficiency consistently below the mean and sometimes well below, indicating that Ryan's aural understanding never moved above a CEFR A1 level. The correlation coefficient between listening proficiency and time was 0.18, indicating a weak relationship between time and progress. Ryan's Time 1 and Time 12 scores were among the seven lowest scores achieved within the entire dataset (1452 data points), and his highest score, achieved at Time 7, was in position 719 of the rank order. (Contrast this with Theo, who occupied six of the highest ten places overall.)

Table 52: Ryan's listening proficiency, compared with the mean

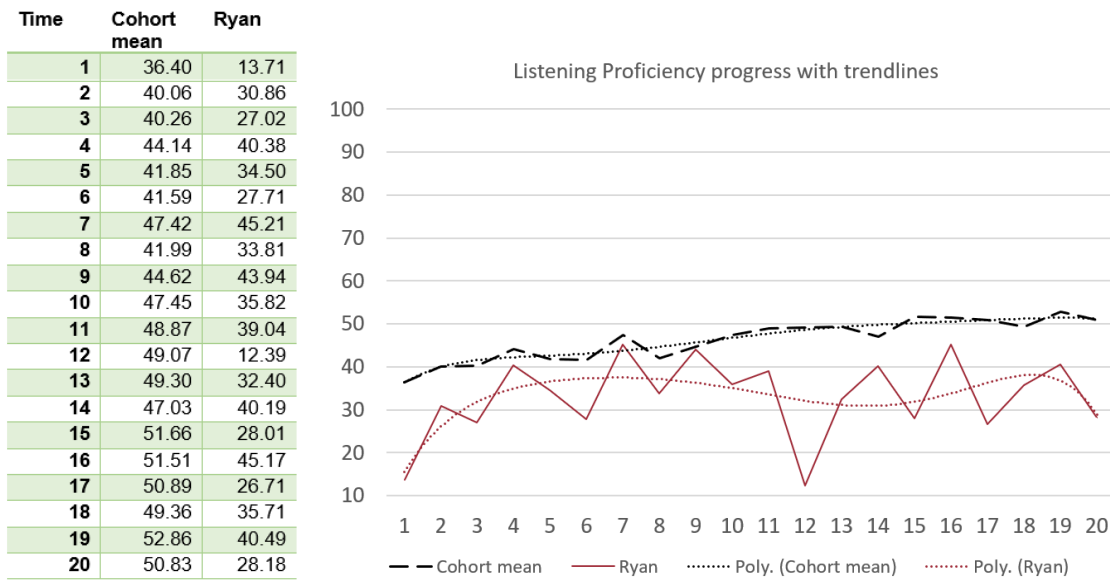


Figure 100: Ryan's listening proficiency, compared to the mean, with trendlines

A moving min-max graph can be seen in Figure 101. The maximum line remained fairly level throughout the study period: we can see a widening of the bandwidth in the middle of the study reflecting his particularly low score at Time 12. The final part of the graph from Time 15 onwards is very similar to the window between Time 5 and Time 8.

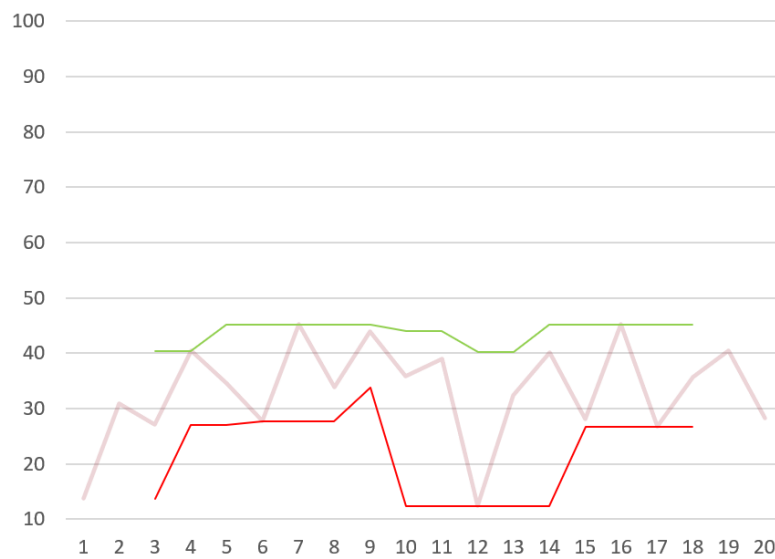


Figure 101: Ryan's listening proficiency moving min-max

The change point analysis tool indicates that Ryan’s listening proficiency remained in the same phase throughout the three years of the study; no second phase of proficiency was detected within the data. The change point graph is illustrated in Figure 102.

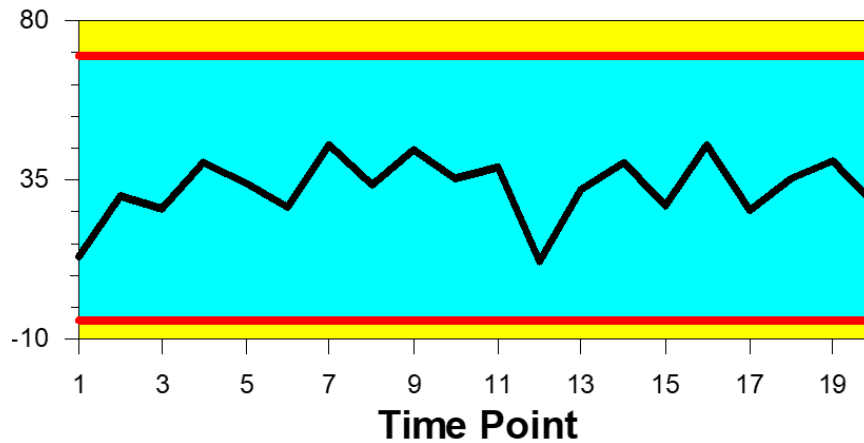


Figure 102: Ryan’s change point analysis: no change

9.3. Ryan’s linguistic subsystems, their development, and relationships

9.3.1. AuralLex scores

Ryan’s vocabulary scores as measured by the AuralLex test are given in Figure 103. The correlation coefficient between vocabulary size and time was 0.14, indicating a weak relationship between time and progress. The graph shows several large peaks, in particular Time 8 where Ryan’s vocabulary size was estimated at 900 words, but in the main his trendline stayed well below that of the cohort mean. Highpoints at Times 17 and 18 were indicative of potential learning, but his subsequent scores did not maintain this progress. The four tests of equal difficulty rotated throughout the study and are colour coded in the x axis for transparency; specific tests were not the cause of the fluctuating scores.

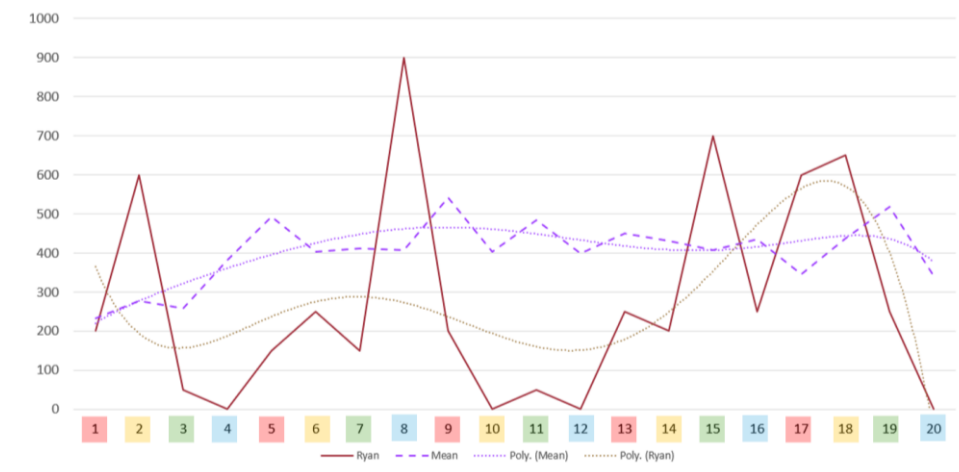


Figure 103: Ryan's vocabulary size compared with the mean

Ryan's vocabulary score as moving min-max is given in Figure 104. The moving minimum score rarely exceeded 0, although there is much more variation in the moving maximum score; in other words, estimations of Ryan's vocabulary size were erratic throughout the study.

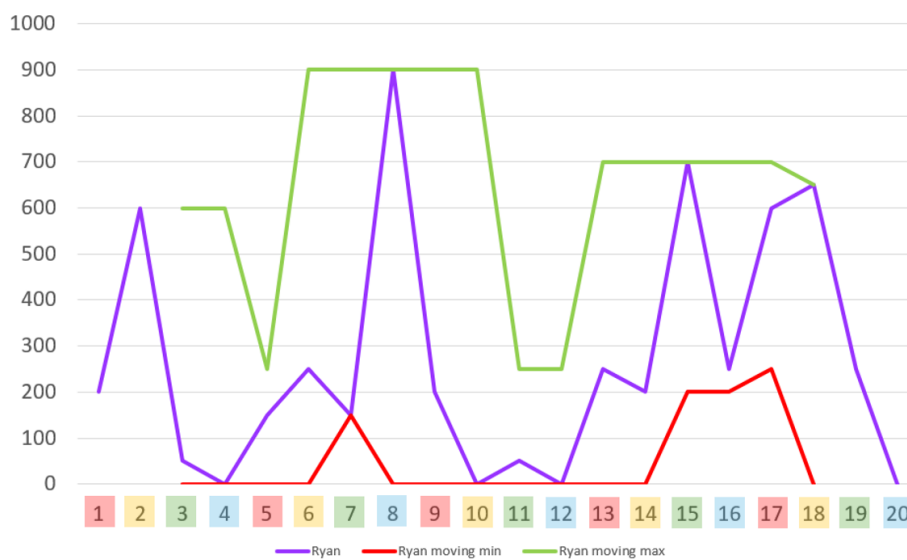


Figure 104: Ryan's AuralLex moving min-max

Ryan's small vocabulary and little evidence of progress can also be seen through his paused translation offerings and discussions of these in our interviews. For example, at Time 2 he managed to translate *je n'aime pas* ('I don't like'), *j'aime* ('I like'), *je préfère*

(‘I prefer’) *piano* (‘piano’) and *guitare* (‘guitar’). He did not manage *j’adore* (‘I love’).

At Time 3 his only accurate attempts at translation were *je m’appelle* (‘my name is’) and *je préfère* (‘I prefer’).

Time 6 took place at the end of the first academic year of learning French, at which time Ryan was not able to give me the meanings of the words *je* (‘I’), *et* (‘and’) or *avec* (‘with’). By Time 8, he had added *beaucoup* (‘a lot’) to the list of words he could recognise at Time 2. Indeed, at Time 15, after two years and two months’ tuition (albeit with a covid lockdown), he was unable to tell me at interview the meaning of *un* (‘one / a’) or *onze ans* (‘eleven years’), and at Time 18, when presented with *mes amis* (‘my friends’), he told me,

‘I know it, but I can’t place it. It’s like we’ve revised the difficult things, but we haven’t revised the simple things we might need.’

Ryan ended the study still unable to translate *je* (‘I’), the implications of which will be addressed in chapters 10 and 11.

9.3.2. AuraGram scores

Ryan’s grammatical development is shown in Figure 105. (The six tests of equal difficulty are colour coded in the x axis.) While Ryan’s score was well below average until Time 7, from Time 8 to the end of the study he scored broadly in line with the cohort average, and the correlation between time and score was strong at 0.541.

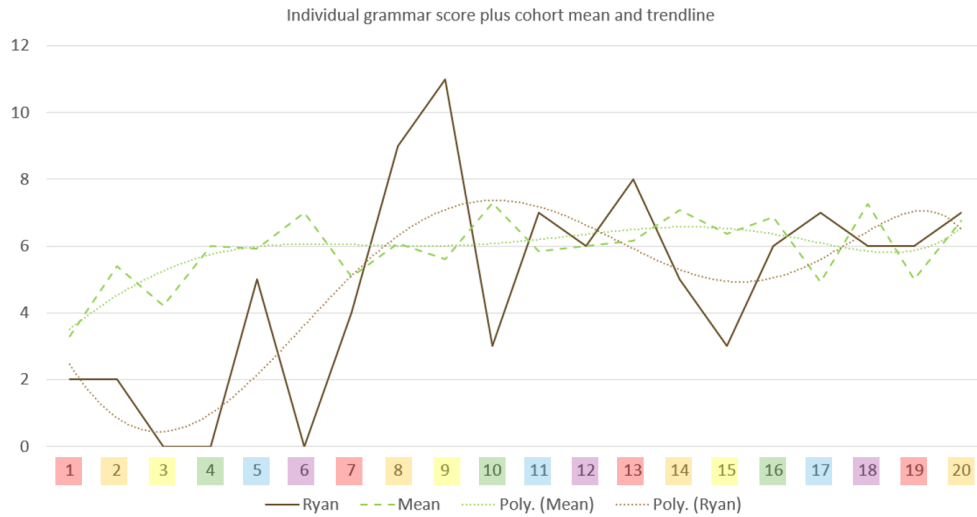


Figure 105: Ryan's AuraGram scores compared with the mean

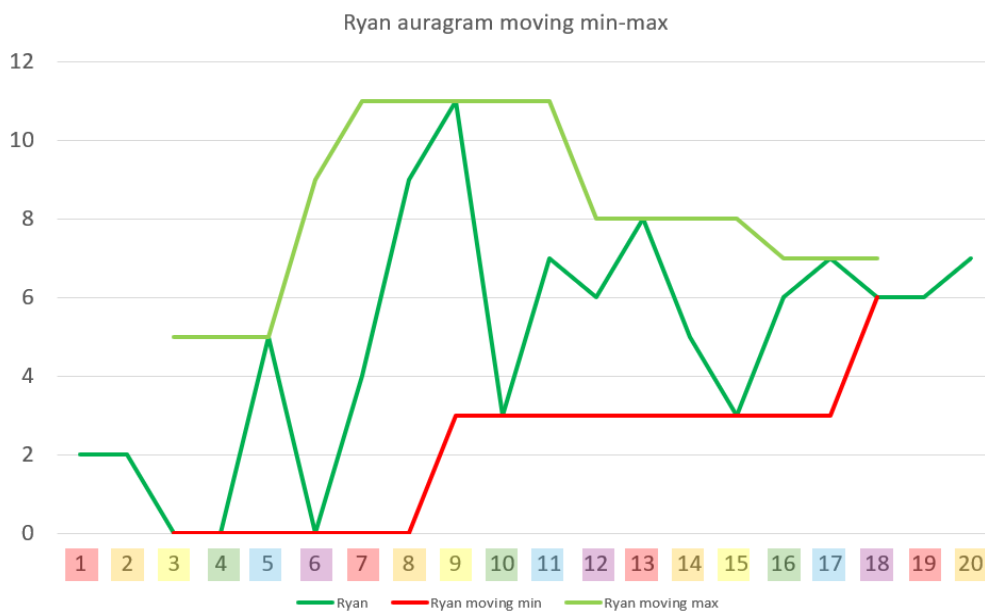


Figure 106: Ryan's AuraGram moving min-max

Ryan's moving min-max scores are shown in Figure 106, highlighting fluctuations in bandwidth but with clear visible increase in the minimum score, albeit with a less predictable trajectory in the maximum score. The bandwidth around Times 7 to 9 shows a

wide fluctuation at that time, and the subsequent narrowing suggests an inclination towards a more stable system.

Ryan was quite preoccupied with the AuraGram, and he raised the topic at every interview until Time 6. In early interviews he told me that, ‘*I put I don’t know because I didn’t know what to do.*’; he struggled initially with the very concept of a grammaticality judgement test. At Time 19 when asked to reflect on being a case study, he again discussed how hard the GJT was.

9.3.3. Developing relationships between linguistic subsystems

In order to address the relative roles of lexical and grammatical knowledge in Ryan’s developmental trajectory of listening, Figure 107 places his vocabulary, grammar and listening scores on the same graph, normalised on a scale of 0 to 1. As stated in section 5.2.1 of Findings 2, the 0.2 line represents a vocabulary size of 175 words. In grammar terms, the maximum observed score within the whole cohort was 15/20, so the 0.45 line represents a score of 7/20.

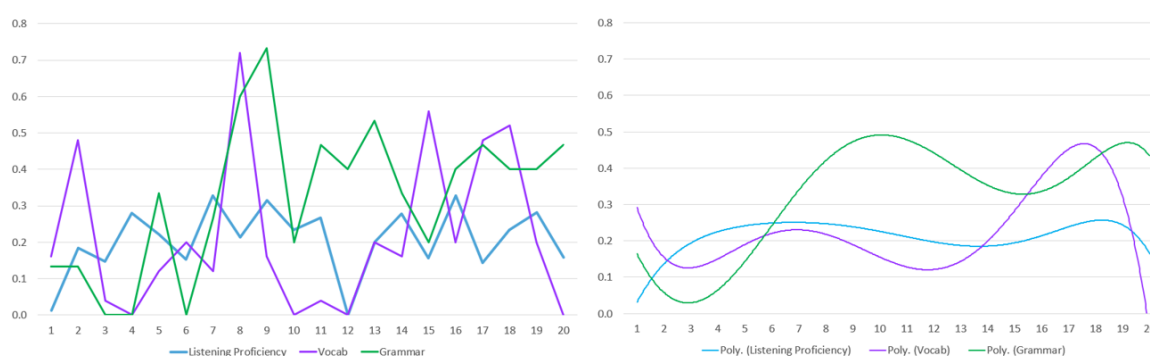


Figure 107: Ryan’s listening, vocabulary, and grammar on a single chart: raw scores and trendlines

The trendlines indicate a similar trajectory for vocabulary (purple) and grammatical (green) progress until Time 5, at which point grammatical proficiency increased.

Between Time 3 and Time 14, vocabulary size and listening proficiency trendlines were

roughly in alignment. It is noteworthy that Ryan’s grammar scores were higher than his vocabulary and listening scores, and this suggests that he was not able to access the grammatical knowledge measured in the AuraGram test in order to help him with his listening.

Ryan’s moving correlation scores between listening proficiency and grammar and vocabulary are shown in Figure 108. They suggest a competitive relationship initially between vocabulary and grammar, a central phase in which both subsystems correlate strongly positively with listening proficiency, before reverting to a competitive relationship for the final phase.

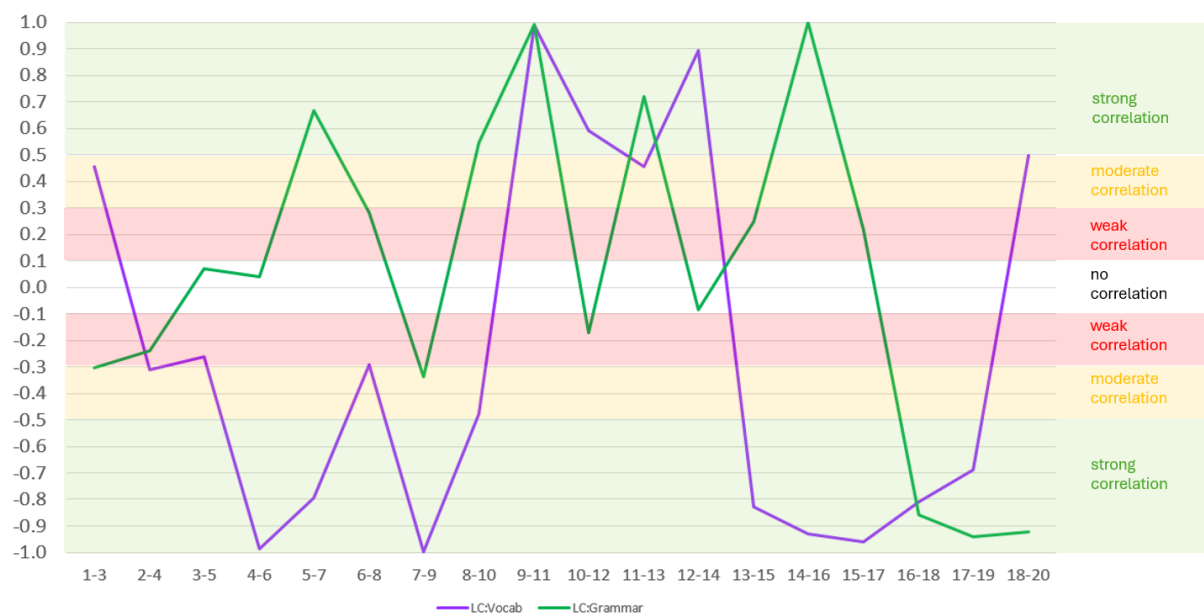


Figure 108: Ryan’s Moving correlations between listening proficiency and vocabulary / grammar

9.4. Ryan’s listening errors and processes

Ryan’s account of his experience of French lessons paints a picture of a student frequently in a state of confusion both about the immediate listening tasks at hand, and also about the broader pedagogy to which he was exposed.

In terms of listening, at the outset (Time 1) he identified the issue of unsatisfactory feedback when reviewing a recently completed task, saying, *‘the teacher only says it once, and later in the lesson they tell you what it was, but you don’t remember like the actual sentence,’* pinpointing a futility of the comprehension approach, particularly for a student with poor working memory, and also suggesting an early unmet desire to achieve. At Time 2 he told me, *‘It’s like a mouthful of French, all in one go, just a really long sentence. It’s all a bit too quick,’* and the image of the ‘mouthful’ returned often when he could not segment speech.

By Time 6 he appeared to be struggling when he told me,

‘French listening is difficult because you have to listen to it, not look at it. It’s a big chunk of writing. It’s kind of like, bla bla, bla bla, blu blu blu bla. In my head. In a way.’

Nonetheless, like Theo, he valued the quiet time in class afforded by listening, although he made ongoing references to ‘gobbledegook’ throughout the study.

9.4.1. Understanding dominated by cognates

Within Ryan’s entire paused translation dataset, and triangulated against that of the stimulated recall, almost all correct translations were cognates or pseudocognates – key exceptions being *j’aime, je n’aime pas* and *beaucoup*. He appeared to make some progress in his auditory recognition of cognates, although it was not consistent and it is sometimes surprising both what he managed and failed to translate at various times. I have already discussed in section 9.3.1 his reliable understanding of the cognates *préfère* and *déteste*, although, like many other participants, he tended to conflate *préfère* and *préféré* (‘favourite’).

Pseudocognates appeared in Ryan’s dataset from Time 10 onwards, such as ‘music’ for *amuser* and ‘silver’ for the name *Sylvie*. These were inconsistent – for example he could not translate *gentille*, /ʒãti/ although he successfully managed *égoïste* /egɔist/ (‘selfish’). At Time 14, he told me, ‘*I’ve heard cousine before, but I can’t piece it together. Let me think. I just can’t think of it.*’ I discussed the ‘sound trumps sense’ hypothesis in the main cohort data and this was definitely true of Ryan’s experience, with such examples as ‘clay’ for *sucre* /sykœ/ at Time 16 and ‘vaccinate’ for *vacances* /vakãs/ at Time 17 (December 2020, just as the UK covid vaccination programmeme was beginning).

Towards the end of the study, Ryan gave three mistranslations in our stimulated recall suggestive of wider lexical knowledge. These again point to a qualitative level of progress made even if this is not apparent in the quantitative data. All of these are French-French pseudocognate traps: both *campagne* and *camping* translated as ‘friends’; *chambre* as ‘sing’ (confused with *chanter*) and *seul* as ‘sun’ (confused with *soleil*).

9.4.2. Words but rarely sentences

Not only was Ryan’s understanding dominated by the recognition of cognates, but throughout the study it remained firmly on the level of individual words; at Time 3 he told me, ‘*I can pick out a few words, but I can’t do the whole sentence,*’ and his account at Time 20 was:

‘I’m hearing the words which I know and they kind of like float around. They float around in French. And then as I go, I take words out and put them in English.’

Even at the word level, recognition and understanding was inconsistent. For example, in response to the phrase *J'adore mon cheval et je fais souvent de l'équitation*⁶⁹ ('I love my horse and I often go riding'), he told me, 'I don't know any of those words,' although in his previous utterance he had said that 'swimming and horseriding look kind of alike'. Above I referred to a conflation of *préfère* and *préféréré*; at Time 4 it appeared that Ryan had also misappropriated the word *beaucoup* and tended to translate any verbal phrase qualified with *beaucoup* as 'I quite like' (from *j'aime beaucoup*, which he had confused with *j'aime assez*).

In section 4.3.3 of Findings 1 I discussed the phenomenon of 'doubling up', whereby a listener would translate a single phrase twice (e.g. *on est allé* became 'we went to Italy'). Ryan did not provide any evidence of this in his paused translation, although he did so twice during stimulated recall (where he often was able to translate far more for me than he had done during the test battery). For the translation of *j'habite dans une petite ville*⁷⁰ ('I live in a small town'), he told me that the sentence meant 'me and my family live in a small town', explaining himself by saying '*petite* means little, and I think the other word was family'. In the same interview he translated *je déteste ça parce que c'est ennuyeux* ('I hate that because it is boring') as 'I hate parkour because it is boring.' In other words, he translated *parce que* twice, once as 'parkour' (making it into a pseudocognate), and then correctly as because. There is no other evidence in the dataset of Ryan doubling up his translations.

⁶⁹ T4 track F9 track 3 (also at T7 and T10)

⁷⁰ Test E56 track 2 T6 (also at T13 T19)

9.5. Summary of findings for Ryan's linguistic systems

Ryan's linguistic developmental trajectory differs from the mean, in that:

- Negligible progress was made in listening proficiency or vocabulary, although some suggestions of progress were detected through the qualitative data.
- Although progress was made in grammatical knowledge, there was no evidence that this contributed to his listening proficiency.
- Ryan's listening proficiency was dominated by single words, most of which were cognates.

9.6. Ryan's metacognitive subsystems, their development and variability

This section will explore Ryan's self-report through the MALQ and where feasible, triangulate it with the qualitative data collected through error analysis and the semi-structured interviews. Ryan's own account frequently goes beyond metacognitive awareness, and contains extensive discussion of his broader behaviour in the classroom as well as constructs of self-efficacy.

9.6.1. Metacognitive scores

Ryan's scores on the Metacognitive Awareness Listening Questionnaire are given below in Figure 109, with moving min-max scores in Figure 110. The statements were rated on a six-point Likert scale ranging from 'strongly agree' to 'strongly disagree', where 'strongly agree' would equate to a score of 1.0.

Figure 109 demonstrates that Ryan's responses were largely positioned between the 'agree a bit' and 'agree quite a lot' and fluctuations in all four subsystems were minor.

Ryan's Person Knowledge⁷¹ scores at the outset suggested that he felt challenged and nervous during listening, but that over time, he felt slightly better, with a notable dip at Time 7 where he reported much lower agreement with these statements. (Recall his personal best listening proficiency score also took place at Time 7). In the second half of the study, his Person Knowledge scores hovered around the middle of the scale. Ryan's response to the Directed Attention⁷² and Problem-Solving⁷³, by contrast, never dipped into the 'disagree' section of the scale, although the moving min-max graph demonstrates wider variability within the Directed Attention construct than for that of Problem-Solving. As far as Mental Translation⁷⁴ was concerned, on two occasions Ryan reported that he was very strongly inclined towards this strategy while listening, although this inclination fell away towards the end of the study and the narrowing bandwidth suggests an approach towards a stability in this subsystem.

⁷¹ listening in French is more difficult than the other skills / listening comprehension is a challenge / I feel nervous when listening

⁷² I focus harder when I have trouble / I recover lost concentration rapidly / I try to get back on track / I don't give up when I have difficulties

⁷³ I use the words I understand to guess the words I don't / I compare what I understand with what I know about the topic / I use my experience and knowledge to help me understand / I adjust my interpretation if I know it's not correct / I use the general idea of the text to help me guess / I check back to see if my guess makes sense

⁷⁴ I translate in my head as I listen / I translate key words, I translate word by word

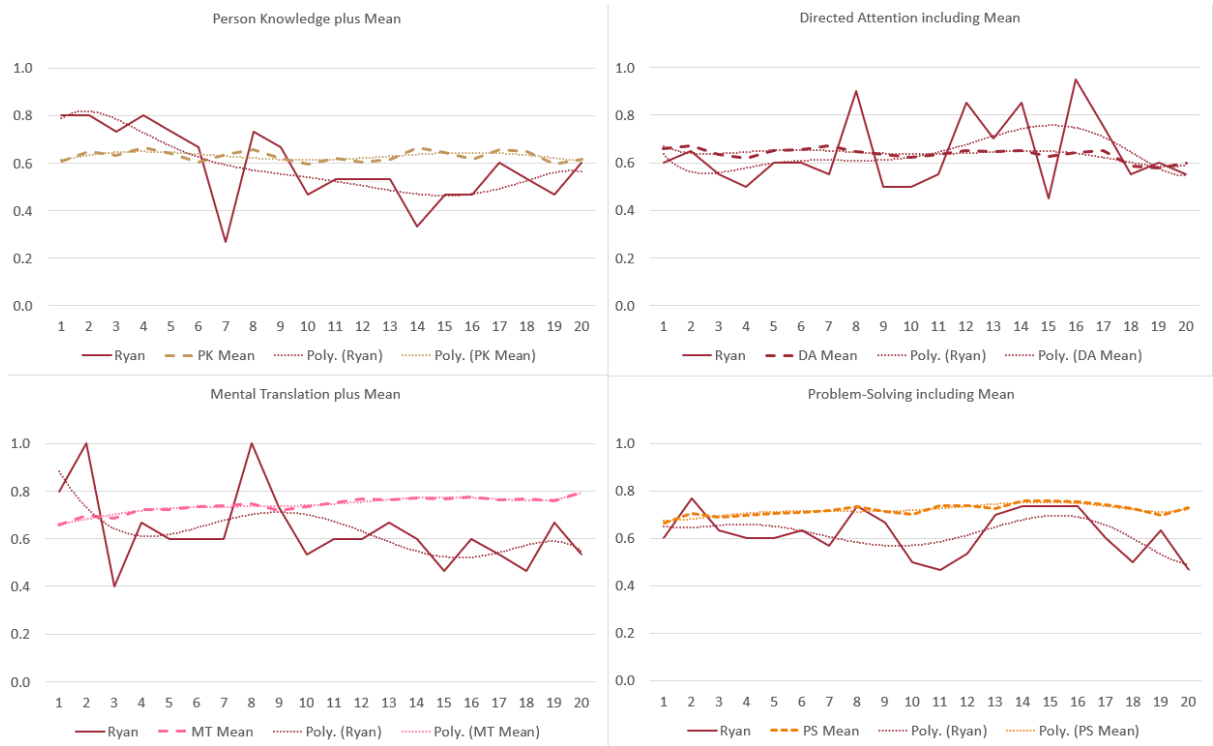


Figure 109: Ryan's metacognitive scores plus cohort mean, with trendlines

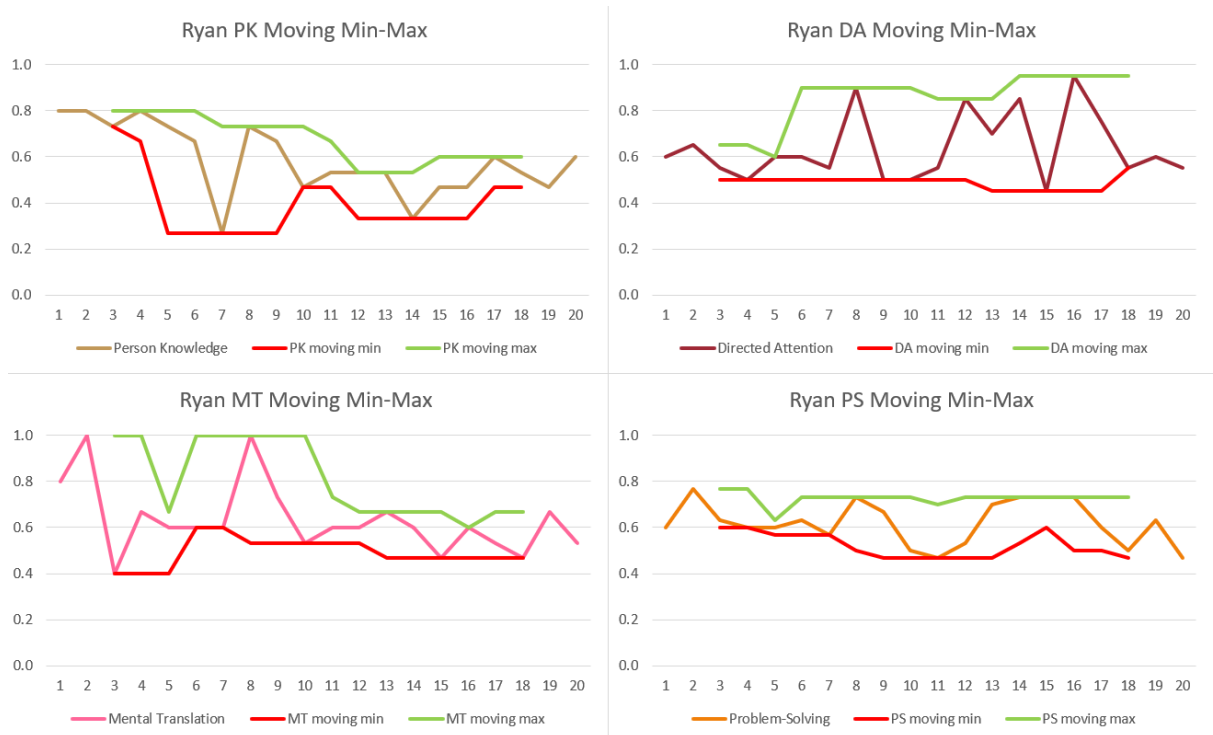


Figure 110: Ryan's moving min-max for the metacognitive subsystems

9.6.2. Developing relationships between metacognitive subsystems

In section 5.3.2 of Findings 2, I showed that on average, trajectories of Mental Translation (pink) and Problem-Solving (orange) followed each other, and that the same was true of Directed Attention (burgundy) and Person Knowledge (beige). The previous case studies have shown that for each individual, a different pattern emerged. In Ryan's case, in Figure 111, Directed Attention and Problem-Solving appeared to align, as did Mental Translation and Person Knowledge, although to a lesser extent. Trendlines suggest that Ryan might have been inclined to hedge somewhat in his metacognitive self-report, given that after Time 2, all trendlines were positioned between 0.45 (very slightly disagree) and 0.8 (agree quite a lot).

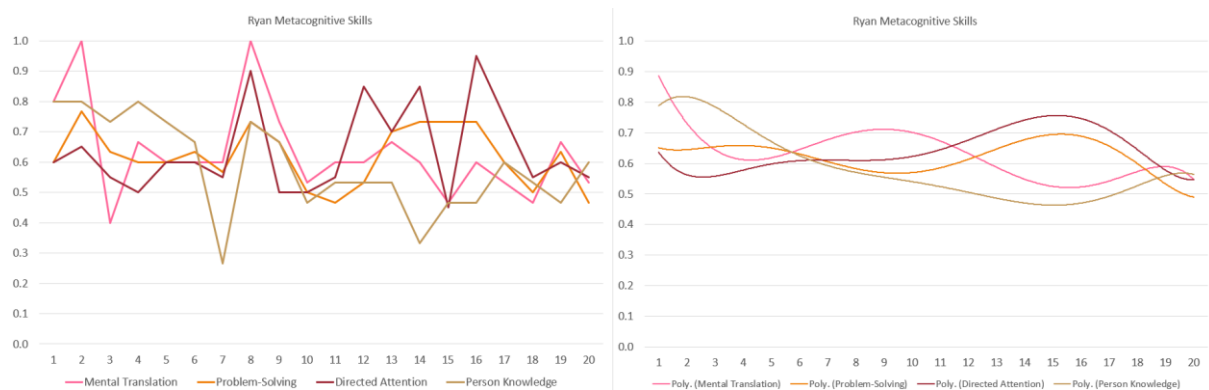


Figure 111: All Ryan's metacognitive subsystems plotted together: raw scores and trendlines

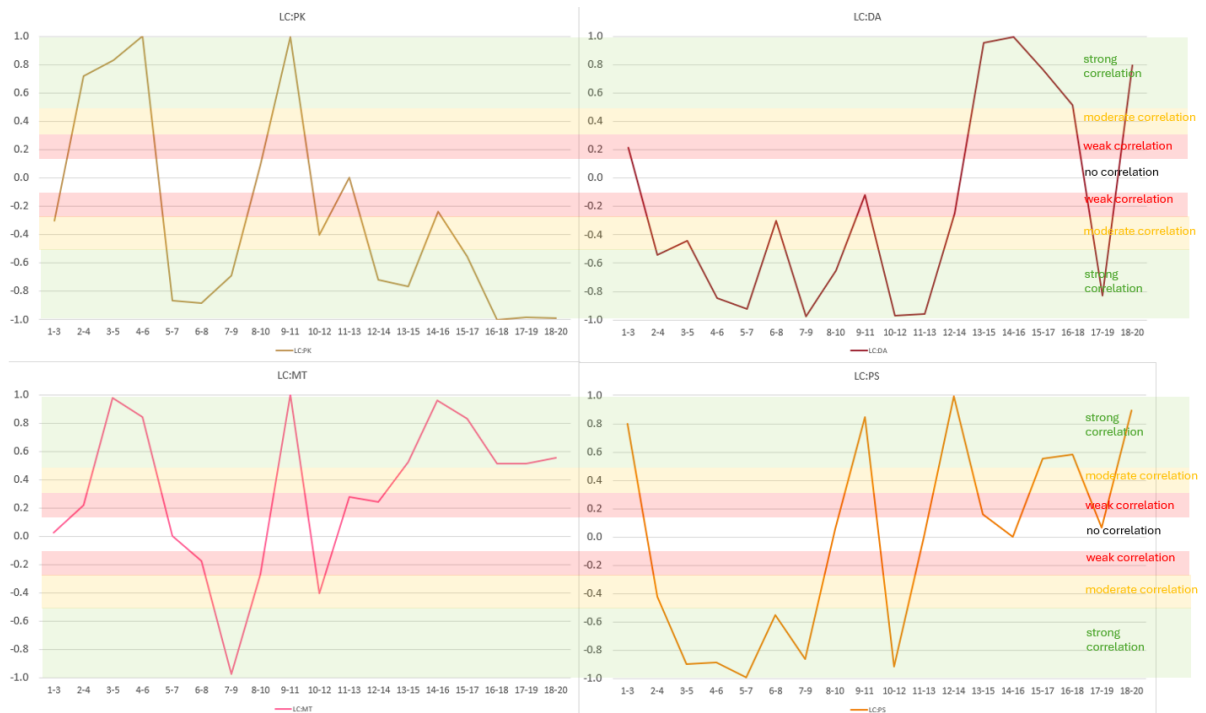


Figure 112: Ryan's moving correlations for the metacognitive subsystems

Figure 112 illustrates the moving correlations between Ryan's listening proficiency and each of the four metacognitive subsystems, although it is important to bear in mind (as was also the case with the other case studies) that the nature of the instrument means that although the moving correlations are shown to be strong, many of the underlying fluctuations within the metacognitive measures were subtle.

Figure 111 shows the proximity of the trendlines for Directed Attention and Problem-Solving, as well as those for Mental Translation and Person Knowledge, Figure 113 also overlays their moving correlations. It demonstrates a close relationship between Mental Translation and Person Knowledge throughout the first half of the study (i.e., 'if I am anxious, I translate') which changed to one of competition in the second half ('if I am anxious, I do not translate'). The close relationship between Problem-Solving and Directed Attention became only briefly competitive in the third quarter, and then returned

to something akin to a supportive relationship (‘if I pay more attention, I am more likely to use problem-solving strategies’).

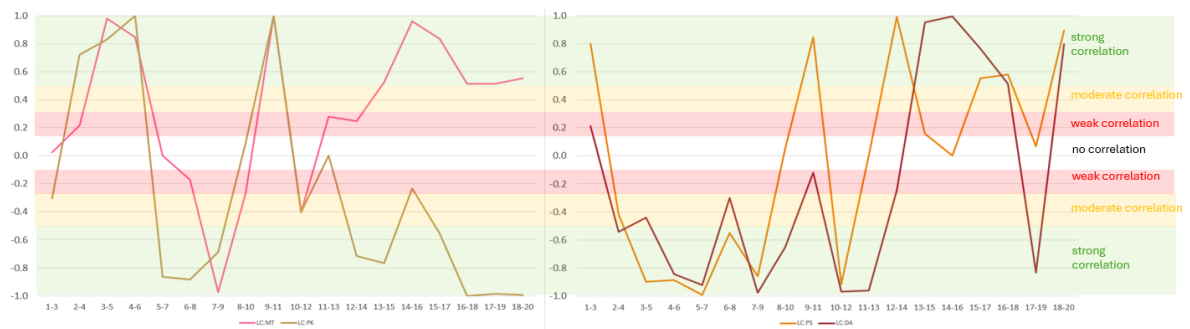


Figure 113: Ryan’s Metacognitive subsystems moving correlations, overlaying Mental Translation with Person Knowledge, and Problem-Solving with Directed Attention

9.7. Ryan’s strategic journey

This section addresses the non-linguistic findings borne out of my semi-structured interviews with Ryan over the three years. I use the term ‘strategic’ loosely: here I will tackle Ryan’s broad view of schooling and his approach to life, my own struggles to bring sense to what he told me, and divide his journey into three phases based on the frequency of various themes that he brought up at interview.

9.7.1. ‘I didn’t mean to hit Sir’: An impulsive sense of fun

I referred in section 9.1 to Ryan’s outgoing nature, which he recognised impacted both his learning and his achievement in my test battery. He told me at Time 4 that he had made the decision to place himself in the corner of the computer room during testing, in order to reduce distraction, yet this did not always remove all sources of extraneous entertainment. At Time 8 another participant ‘*did a huge sneeze,*’ during the testing, which Ryan reported to me afterwards ‘*was really funny; you could hear it half way across the school.*’ Time 19 saw Ryan also distracted by laughing uncontrollably when a

rebellious student was reprimanded for her lack of school uniform while he was working on the test.

In French lessons, Ryan received several formal school punishments due to his poor behaviour, although he did not take responsibility for this, and told me at Time 4 that it was due to other students in class *'who always mess about, and they always dump the stuff on me.'* Yet perhaps he was aware of his own role when he told me that he had missed parents' evening. I asked him what he felt his teacher would have said, to which he answered, *'That I was a well behaved boy but sometimes, when we're doing something over-fun, I over-react.'* Ryan's sense of fun is summed up in a post-hoc account given at the end of the study, in which I asked him to tell me about any memorable episodes in his learning. He discussed an in-class fashion show that was the culmination of a topic on clothing and style:

'Our group, we showed up in a dinosaur costume. I have never forgotten that. We literally had the biggest fun times of our life. We took two whole days to get some paper mâché and build a structure. Helmet, everything. We got a tail, we got wings, we got everything.'

Ryan's behaviour in French was typical of his behaviour across the school. He told me at Time 18 that school was for, *'actual decent fun times,'* and reminisced about a particular incident involving a handful of clay that had taken place during a pottery class in the previous academic year:

'Year 8's supposed to be a fun year, isn't it? I think I could've eased on the fun part of it. I didn't mean to hit Sir. I'm sitting here, at the end of the classroom. And there's a middle table. I meant to chuck it at that, but I had really sticky fingers because we were doing clay. It slipped before I could

gently underarm throw. So somehow it bounces off Jacob, then it hits Joe, and then it kind of like slides across his shoulder blade. And then it just rolls onto Sir's back. I'm like, "I'm sorry", but he's just like, "You shouldn't be throwing it anyway."'

This anecdote exemplifies not only a sense of fun, but also an inclination towards impulsiveness and a lack of ability to see a broader picture or the wider implications of his actions; notions that can perhaps be carried over into both his learning behaviours and his comments to me during interview.

9.7.2. An inconsistent message

It could be said that the only consistent theme that can be drawn from Ryan's interview data is its inconsistency. In my initial analysis I drew out contradictory themes of strong self-efficacy yet also of defeatism; of a 'can-do' strategic Problem-Solving laced with reports of 'gobbledegook'. To illustrate the most accurate picture of Ryan's strategic developmental trajectory is not to divide it into themes, but to examine it chronologically. This will acknowledge and the repeated contradictions while allowing me to explore complex dynamic systems that might emerge. These themes are:

- A strong sense of self-efficacy
- A can-do attitude in the face of uncertainty
- Insecurity about a task
- A Problem-Solving approach: using context to compensate for incomplete understanding
- Daydreaming
- Defeatism
- Mind going blank

Table 53: Ryan's discussion of approaches to listening over time

Time	Self-efficacy	Can-do attitude	Insecurity	Problem-Solving	Daydreaming	Defeatism	Mind going blank
1							
2				✓		✓	
3				✓	✓	✓	
4	✓						
5			✓		✓	✓	
6		✓		✓		✓	✓
7	✓					✓	
8	✓		✓				✓
9	✓			✓		✓	✓
10	✓	✓					✓
11	✓				✓	✓	
12		✓					✓
13						✓	
14							
15			✓				
16		✓					
17							
18		✓				✓	✓
19	✓	✓			✓	✓	✓
20		✓					

Table 53 illustrates the location throughout the study of when Ryan discussed each theme with me, and allowed me to split his trajectory into four phases, which will now be discussed.

9.7.3. Times 1-6: 'I don't know what anything means, but I kind of wanted to do it'

Ryan referred frequently in our interviews to his tendency to daydream. At Time 3, he said, 'I get kind of daydreamy.' During the Time 5 test battery, having been able to get to know him through preceding interviews, I watched Ryan complete the test; during this time he frequently moved the mouse around seemingly at random and was not systematic about his approach. Asked about it afterwards, again he told me was daydreaming, adding, 'I do that a lot.' He seemed resigned to his lack of achievement; he told me that

he was not frustrated by that although he agreed that the test battery had been hard; in fact he said,

'I don't really think about it that much. I just move on. I think about it for about five seconds and then I'm like nope, nope, nope.'

Still, by Time 5 he *did* express frustration in his sense that *'I don't know what to do but everyone else does,'* yet also alluded to a deeper irritation of being held to higher expectations rather than being allowed to give up, when he told me,

'If I tell someone I got it wrong, something will happen. I'll get talked to, and I'll have to do another one, like 500 times or something.'

This suggests that even aged 11 Ryan had developed a distrust of disclosing his academic struggles due to a dislike of the interventions that he expected would ensue.

A classic compensation for incomplete understanding is to begin with what one does understand, and then to use a range of strategies to fill in the gaps. Helen, Theo and Fran all used this approach to varying effect. Ryan also frequently reported using this strategy; more so towards the end of this first phase, telling me at Time 6,

'There's most probably going to be a few words I know, so I put the words in. Then I remember what they said before. And I put the other bits which would make sense.'

Still, this self-report was not entirely accurate: during this phase, Ryan's sense of context was not always appropriate. For example at Time 3 he was asked to translate a short exchange in which the speaker gave their name and then spelled it⁷⁵. His mistranslation

⁷⁵ Time 3 test G78 track 2

did indeed attempt to build meaning from what he had understood, but it was *'I know my vowels. A E I O U.'*

More frequently, Ryan did not seize on the words he knew and then *'put in other bits that would make sense'*. This can be seen both at Time 5, where he heard *je ne mange jamais de poisson. Beurk!* ('I never eat fish. Yuck!')⁷⁶ – a track that contained an audible tone of revulsion. He wrote nothing, and his account of this at interview speaks to his confusion:

'I just didn't... at the time I kind of – not nervous – but panicking what to put because I didn't know any of it, but I kind of knew what it was saying.'

At Time 6, Ryan began to discuss the difference between my data collection, which he found difficult, and listening tasks in class, which he reported to be very easy; a theme that would come up again much later. His approach in class was confident, telling me that in classroom listening tasks he *'knew all the words, and when I get them right I feel kind of proud of myself.'*

By contrast, during the Time 6 test battery he struggled. Although he reported that he never gave up, that he wanted to learn and that he always concentrated hard, in response to *j'habite dans une station de ski à la montagne* ('I live in a ski resort in the mountains'), his concentration and perseverance was misplaced. He had failed to recognise *ski* and told me:

'I didn't write anything because I didn't know what montagne is... If one word really puzzles me, it kind of makes me off track to the rest of it.'

⁷⁶ Time 5 test F28 track 5

In other words, he had identified that a key word was *montagne* ('mountains'), but became hyperfocussed on it, to the detriment of a more holistic understanding.

Only two tracks later he heard *j'habite avec ma mère* and this time he wrote 'I live near the seaside', correctly translating *j'habite* ('I live') and identifying the homophonous word *mer* (seaside) for *mère* (mother). For Ryan this was a successful translation, and I asked him about his feelings:

'I feel like I'm surprised, because most of the time I think there's no point in doing it because I don't know what anything means. But I kind of wanted to do it.'

This growing can-do approach is a departure from the workshy attitude only six weeks before at Time 5, and might pave the way for the new, although not always founded, sense of self-efficacy which will be explored in section 9.7.4.

9.7.4. Times 7-11: 'I've always got them all right' but 'it just poofs out of my mind'

Ryan's second phase within his cognitive and metacognitive trajectory was characterised by a strong sense of his own brilliance, oblivious to the contradictions he was making when he would also tell me – for example at Time 7 – that *'I didn't understand pretty much any of it,'* at Time 9 that what he had heard was *'gobbledegook'*, and at Time 11, *'if I focus on one thing, I forget about the other thing. It just poofs out of my mind.'*

At the same time as references to a lack of understanding and a blank mind (with reference to my test battery) when discussing listening in class, Ryan and I had the following conversation at Time 8:

Let's imagine Mrs A has said 'OK, we're going to do a listening now', we'll look at page 12 exercise 2. And then she's about to press play on the machine...

I'm... when I look at the page before she presses it, I already know the answers, but in the listening survey it's a bit different, because we... cos you're not taking it from a book.

So how do you already know the answers before she's played the track?

Because the first time I look at the page it just comes into my head. Well most of the time I know them.

So you sort of make up the answers?

Yeah. Most of the time they're correct. For some weird, weird reason.

OK. So do you have any special strategies that you use when you're doing that sort of listening in class? Any tips and tricks to make it easier?

Um, I just shut out from everyone around me except for my partner, and I just focus. Really hard. And I've never missed one wrong. All the time we've done listening in year 7 and year 8. I've always got them all right. Not one wrong. Except for the very first time we had listening in year 7.

Ryan was never again quite so expansive about his abilities, but he told me at Time 9 that he felt he was among the most able five learners in the class. At Time 10 this was further qualified by his statement: *'I don't think definitely I'm the best, but I think in my mind I'm the best.'*

It is possible that this self-efficacy had a positive effect on his ability in my test batteries, which were harder than the Comprehension Approach style exercises in the classroom. His stimulated-recall struggles with a Time 9 track indicated the swings between self-efficacy and defeatism that were taking place seemingly on a moment-to-moment time scale:

Yeah. I was so confused. Like – why?

Why what?

Why can't I hear anything.

And what happens when you feel confused and thinking why can't you hear anything?

Um – I just shut off, really, and I just don't think. But I just realised I know a few of those words, and, um, it just popped into my mind that I could actually do it.

Similarly, Ryan talked at Time 10 of the importance of attempting a task even if he lacked confidence in the result, and at Time 11 discussed the need to research the meanings of unfamiliar vocabulary, suggesting 'asking Miss, or looking in a book, my student booklet or Duolingo,' as sources of definitions. When asked whether he could remember the unfamiliar vocabulary he needed to look up, he said he did not. In other words, Ryan was able to describe to me the behaviour of a motivated student, yet was not able to implement it sufficiently to meet his own academic needs.

9.7.5. Times 12 to 17: 'I searched up how to be a better rememberer'

The period between Time 12 and Time 17 included two periods of school closure, although Ryan's additional needs meant that he had returned to school before Time 13, and was entitled to attend school in the January to March 2021 lockdown. Nonetheless, lessons were still being taught online, so in practice, attending school meant coming to the building, visiting a computer room which was supervised by a rota of teachers, and participating in the same online learning as the locked down students.

This phase was characterised by the defeatism seen earlier, not tempered by the same perception of his own ability. He found the Time 12 test battery particularly challenging, but was able to create positive narrative surrounding this, accounting for his difficulty

purely because he had not attended the online lesson immediately before he took the test. Still, Ryan was self-aware enough to acknowledge his memory problems when he told me that he had gone ‘on YouTube and I searched up how to be a better rememberer,’ which he told me he had found very helpful.

By Time 13 he sounded gloomy, saying, ‘*it gets harder every listening that I do,*’ and talked of feeling troubled for the rest of the day if he had struggled with a sentence. And at Time 15 – school had returned and this was the second year 9 time point – he went so far as to say he had ‘lost his touch’, recognising that others in the class had both better memory and better knowledge. He retained a level of positivity, though, speaking warmly of his enjoyment of the French atmosphere.

9.7.6. Times 18 to 20: ‘I knew it was wrong, but I wanted to put something’

In the final phase of Ryan’s cognitive and metacognitive trajectory, school had reopened, and Ryan was in a new teaching group, sheltered from the bullying behaviour that had troubled him from Times 14 to 17. As in the previous examples, he recognised the ease with which he completed in-class listening tasks, compared with the difficulty of the paused translation tests, in which he was resigned to underperformance:

‘My mind just collapses when I don’t know it. I don’t mind the fact that I can’t know it, or know what it means.’

Time 19 showed some positivity though: in class he could ‘*answer without hesitation most of the time,*’ and even within the test battery was willing to attempt a response while feeling unsure of himself, but acknowledged that he still ‘*zoned out a tiny bit*’.

In our final interview, Ryan spoke with insightful maturity about his hopes for a military career despite there remaining two more years at school, and remained typically sanguine

about his up-and-down trajectory. He told me he had been removed from class only the previous lesson for bad behaviour, but still concluded our interview with positivity:

‘A lot of people do need to realise that French is a very good subject to learn.’

9.8. Ryan’s metacognitive subsystems: Concluding remarks

My interviews with Ryan frequently surpassed the scope of metacognition and moved onto constructs of self-efficacy, motivation and even memory. One can conclude that:

- MALQ scores suggest that he became less anxious with time, although we saw defeatism in the test battery throughout.
- He reported reduced levels of mental translation, but did not allude to this in interview.
- He reported a boost at Times 13 to 17 in his Directed Attention, but qualitative data seem to contradict this.
- Ryan’s approach and view of in-class listening tasks contrasted strongly with the data provided by my test battery and interviews. In class he saw himself as able and competent, but struggled with the participation in the study.

9.9. Ryan’s complex dynamics: Conclusion

Ryan’s developmental trajectory stands in stark contrast to the other three. Even by the end of the study, his listening proficiency was driven by single words, sometimes reinforced with judicious (or less judicious) attempts to compensate for understanding with (educated) guesswork.

At interview he gave the impression of an erratic learner and listener – someone who wanted to achieve, but also wanted to have fun, someone who had great faith in their

ability while acknowledging that what they heard was ‘*gobbledegook*’. But what is striking is Ryan’s ongoing positivity about his learning even in the face of adversity. His experience provides much pause for thought will be discussed in section 10.3.6, with implications in sections 11.4.5 and 11.5.3.

10. Discussion and analysis

10.1. Introduction

This study examined listening difficulties and the influence of key subsystems both at a moment-to-moment level and over a three-year period, from a CDST perspective. The project aimed to help answer the questions as to the nature of errors made when listening, the relative roles of the linguistic and metacognitive systems which underpin listening proficiency, and the extent to which individuals' trajectories differed from those of the group. It has begun to clarify the opacity of the listening experience, the interconnectedness of listening processes and factors, and the divergence between proficiencies at the group level and individual.

This chapter will discuss and analyse the findings, and it will concern itself with discussion around change over time in its different guises:

1. The changing relationships between listening proficiency and both the linguistic and metacognitive subsystems, at group level.
2. The changing relationships between listening proficiency and both sets of subsystems at the individual level of the case studies.

Given the setting of the research, I will also discuss the unexpected findings and implications in relation to the widening proficiency gap between learners, learners with additional educational needs, and remote learning due to the pandemic.

10.2. Group level subsystems over time

In this section will address my findings with regard to listening proficiency and the linguistic and metacognitive subsystems at group level, contextualised and analysed with current literature.

10.2.1. Linguistic subsystems at group level

Figure 114 (which is a duplicate of Figure 52) demonstrates that when all eighty learners are taken together, there was a consistently moderate relationship between the participants' vocabulary knowledge and their listening proficiency. Furthermore, throughout most of the research period there was also a weak relationship between participants' grammatical knowledge and their listening proficiency.

As I discussed in chapter 6 (Findings 2), there were periods where the lines diverged, hinting at the possibility of competition between vocabulary and grammar, and times when the lines became parallel, which suggested that the two subsystems might be working in consort during those periods (Verspoor & van Dijk, 2011).

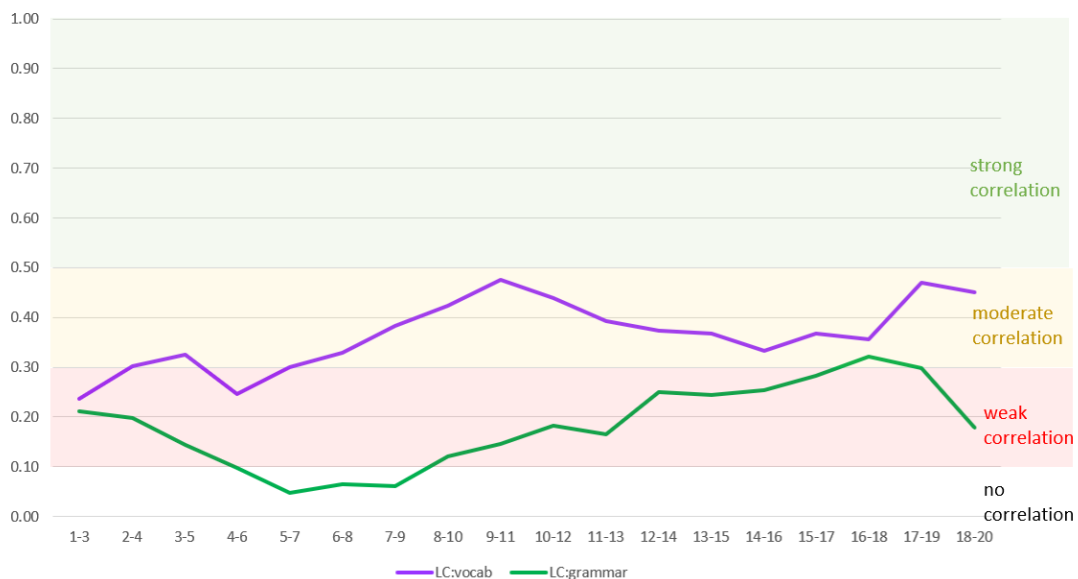


Figure 114: Moving correlations between linguistic proficiency and vocabulary and grammar

To the best of my knowledge there is no previously published research which illustrates longitudinal change in either grammatical and lexical knowledge in relation to listening proficiency, and none which attempts to probe the interactions between more than one subsystem.

Nonetheless, there are two papers which set out a complex, dynamic relationship between vocabulary and grammar. Van Geert (1991) showed that in the first language, vocabulary was a precursor to grammar, but that once a minimal vocabulary size was attained, the relationship changed and became dynamic, eventually settling in a supportive relationship. Verspoor, Lowie, & Van Dijk (2008) made similar findings when researching written second language learning.

Given the lack of research into relationship of both with listening proficiency, in order to situate the present study within the literature, we must revert to cross-sectional research. It is not possible to infer an impression of change over time through cross-sectional research, unfortunately, due to the variance in participants, proficiency and maturity of the different participants.

10.2.2. Relationship between listening proficiency and vocabulary

Despite the fact that it would not be academically honest to attempt to infer change over time from ‘snapshot’ papers, as far as the relationship between listening proficiency and vocabulary knowledge is concerned, my data appear to reinforce the wide body of literature which asserts the close relationship between vocabulary knowledge and listening proficiency. The findings are consistent with those cited in section 2.4.1 of the Literature Review: vocabulary had been found to be a key predictor of listening success among both beginner learners (Vandergrift & Baker, 2015; Matthews, 2018; Smith, 2020) and very proficient listeners (Stæhr, 2009). I also showed that while listening,

participants demonstrated broad difficulties with L2 lexical retrieval – in other words, they struggled to identify words which, given their proficiency, should have come easily to them. Based on the literature, this finding is also not unexpected (e.g. Chen, 2013; Khoshsima et al., 2011; Liu, 2002; Sheppard & Butler, 2017).

Phonolexical errors

In brief, the results associated with phonolexical errors are given below (for a full resumé see section 4.3.6)

- The participants' tendency to identify a 'phantom' English word within a short stream of French.
- A French-English cognate might be a trigger for lexical or segmentation error.
- As a rule, sound overrides sense, with minimal change seen in this over time.
- An increase in the raw number of segmentation errors and vocabulary over time, due at least in part to an increase in attempts at translation.
- An increase over time in the chances of a making a lexical or segmentation error stemming from French vocabulary interference (as opposed to English).

Both the findings set out above and the following discussion situate the problems associated with listening firmly within a perceptual framework (Al-jasser, 2008; Field, 2008b, 2013; Goh, 2000; Tsui & Fullilove, 1998) and speak to the bottom-up element of Field's model of listening (2008b, 2013). Indeed, according to some scholars, these low-level perception problems in listening are the dominant challenges (Goh, 2000; Graham, 2006; Kennedy et al., 2015; Simpson, 2017), particularly in less proficient listeners, as my participants were. As with the quantitative findings discussed and contextualised above, there is no literature which addresses change over time with these errors.

In section 2.2.2 of the Literature Review I discussed the Fuzzy Lexical Representation hypothesis. Cook et al. (2016) describe ‘fuzzy’ or ‘low resolution’ representations of a target word within the learner’s mind, in which the lack of detail when compared with the L1 user can lead to problematic understanding and interference with similar sounding words: this is consistent with the phonological errors that have been found in the present study – for example the *copain campagne camping* confusion, or *vélo ville voile* (see section 4.3.3).

Furthermore, there was only minimal evidence within the dataset that these fuzzy representations improved over time, which might be contrary to the expectation of a typical language teacher (Tsui & Fullilove, 1998). Duyck (2005) showed that L1 influence persists into the L2 even at near-native levels of L2 proficiency, Zhang & Yi Liao (2023) showed that correct differentiation between vowels needed explicit intervention, and Carney (2021) demonstrated that the L1 influence exists not only at word level, but also at phonological level. His L1 Japanese L2 English learners demonstrated phonological interference when differentiating between the words ‘collect’ and ‘correct’; the present dataset shows participants experiencing similar challenges between, for example, the words *tout* /tu/ (‘all’) and *tu* /ty/ (‘you’). Both these examples involve phonemic contrasts which exist in the L2 but not the L1.

I demonstrated in my findings that one aspect of these fuzzy representations was where L2 listeners map words onto sounds asymmetrically – in other words, their translations of some words is consistently correct (e.g. *sœur* /sœʁ/ (‘sister’), or *copain* /kɔpɛ̃ / (‘friend’)), but other, similar-sounding words can also end up with the same (mis)translation. Cutler, Weber, and Otake (2006) made similar findings, whereby their Japanese participants in an eye tracking study were inclined to look at the picture of a locker when presented with the word ‘rocket’, but also looked at the locker when given

the word 'locker'. They hypothesise that the L2 reader has two representations of the two different words, with two separate productive phonetic representations borne out of the reading; but only one receptive phonetic representation, as their aural processing cannot hear the difference. Similarly, Cross (2009) accounts for such asymmetric mapping as a problem in perception of vowel sounds, which I discussed in section 4.3.3. However, I would expand on this hypothesis and add to the perceptual difficulties a complex interplay between the challenges of differentiating between different vowel sounds, and the limited vocabulary of my participants, resulting in their seizing on any word that sounds 'about right'.

There is paltry mention of the misordering of syllables within the L2 listening literature: Wong et al. (2021) recorded 5 out of 640 total errors (0.78%) that were due to metathesis but does not discuss them, and it is alluded to within the work of Rost (2016), and Cook (2012) states that representation of words are not in fixed phonological sequences.

Despite the very small amount of research evidence on misordering syllables when listening, and the relatively small amount of data within the dataset on this, we can be confident that it remains a phenomenon which is occurring within the listener's brain for both my participants and other listeners, and of which those working with L2 listening should be aware. Metathesis is discussed within speech therapy literature on language impairment (e.g. Anzaki, 2021; Leonard, 1982) and there is potential for further research on comparisons between early L2 learning among normally developing adolescents and L1 language impairments which could lead to fruitful pedagogical or therapeutic implications in both fields.

The research cited in the paragraphs above speaks to what I have called the French-to-French problems, in which the fuzzy representation in the listener's head causes them to call up an alternative L2 word. More significant in the present study, however, is the

omnipresence of the L1. Broersma and Cutler (2008) wrote of ‘phantom activation’ of words absent from the input (see section 4.3.1 for examples from my own data, such as ‘Dubai’ for *de bains* or ‘potato’ for *bouteille d’eau*). When L1 vocabulary is activated, Broersma and Cutler explain, the number of lexical competitors for the input significantly increases, which in turn slows word recognition. Triggers for mismatch summarised by McQueen (2007) include incorrect initial phonemes (e.g. honey / money or bun / gun) – rarely seen in the present study; minimal pairs (he cites the English example ‘trusty’ / ‘trustee’), which caused a problem for my participants in an early test with the ages *deux ans*, *dix ans* or *douze ans* /døz_ã, diz_ã, duz_ã/ (‘two years’, ‘ten years’, ‘twelve years’) and when the phonology is consistent with another feasible word (e.g. ‘potato’ for *bouteille d’eau* or ‘Italy’ for *est allé*).

However, Broersma and Cutler (2008) also found in their research that response speed and accuracy increase when a listener is asked to process something a second time in quick succession. My participants were asked to do just this, but Broersma and Cutler’s finding is at odds with the present dataset, as participants continued to experience difficulties despite recency. To contrast with Broersma and Cutler’s work, then, I have shown that only minimal mismatching information can be enough to interfere with adequate comprehension: the low L2 competence levels of my participants, perhaps combined with their youth, would have resulted in ample occasions of mismatch. Broersma and Cutler’s participants were young adults studying at university, and the experiment was carried out with individual words only, which may contribute to the differences with my own findings. These differences also point to the value of research on younger, less proficient learners in better understanding the fuller picture of L2 listening errors.

Textually driven errors

As distinct from the errors discussed in section 4.3, there were also errors which did not appear to be necessarily rooted in the auditory modality of the instrument. In brief (refer to section 4.4.4 for full details), the study found that:

- False friends presented a challenge throughout the data collection period.
- Lexical sets were problematic, although the data made some suggestion that the nature (content versus function words) or type of lexical set might influence how problematic.
- Transfer of training was apparent in the overapplication of ‘I like’ for a variety of other verbs, and this diminished with time.

Each one of these types of error illustrates a situation in which a variety of different words in the participant’s lexicon are competing during the process of making meaning.

To contextualise the false friends issue with that of current scholarship, there is plenty of evidence to suggest that false friends cause problems for even advanced users of an L2 (Dijkstra, 2009). But there is little research on beginners or younger learners. The one exception is Brenders, van Hell, and Dijkstra (2011), who explored the understanding of L2 cognates and false friends in child beginner learners. Their findings – based on laboratory tests with carefully planned lists – were consistent with mine in that false friends were problematic; furthermore, they also reported that where their participants encountered both cognates and false friends together, difficulties were exacerbated.

Recall the tendency of my participants to manufacture pseudocognates, (see section 4.3.1. in Findings 1) and it is clear that coping accurately with false friends is unlikely to be a skill that develops early in L2 acquisition. Couple this with the input-poor environment, and the additional challenges inherent to listening when compared with

reading, and it comes as no surprise that false friends remained problematic even at the end of the study.

Literature concerning the teaching of phonological sets has been conclusive that the teaching of lexical sets is not conducive to learning (Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Nakata & Suzuki, 2019; P. Nation, 2000, 2021; Tagashira, Kida, & Hoshino, 2010): processing takes longer and interference errors are more likely. Yet this approach remains the dominant pedagogy in the UK system (C. Bell et al., 2016), and my findings are to some extent consistent with the literature in suggesting that it is counterproductive, although my data suggest that this also depends on the lexical set itself, with some sets presenting more problems than others. Baxter et al. (2022) suggest an alternative approach revolving around orthographically rather than semantically similar words, however, the data in the present study imply that this might also not be helpful to learners: for example we saw the difficulties in section 4.3.3 whereby pairs such as *vélo* and *violin* were easily confused; and the family words *père*, *mère* and *frère* could be categorised as both semantically and orthographically similar.

To the best of my knowledge there is no literature that discusses progress over time with regard to lexical set proficiency, although Wilcox and Medina (2013), whose intervention study spanned a six week period, hypothesise that learners would make progress, and go so far as to claim that their negative effect for teaching by semantic and phonological groups would have been removed had they been able to carry out a delayed post-test. Their participants were beginner learners of Spanish and consequently have similarities with my own. However, my findings challenge Wilcox and Medina's (2013) hypothesis that progress would have been apparent. Their two lexical sets were low frequency content words (tools and weapons); far lower frequency than time phrases or friends and family words. My explanation for failure to progress in the 'time phrases' lexical set is

rooted in the fact that time phrases are function words compared with the ‘family and friends’ lexical set which are content words (see section 4.4.2), meaning the words are both less perceptually salient and less easy to guess. But given that even successful translation of family and friends words seemed to vary according to the input, we can yet again point to the interconnectedness of various factors (here salience, co-text and context) which are all likely to confound a participant’s successful translation, and suggest that this is an area where progress is particularly slow.

As far as the overapplication of the common phrase ‘I like’ is concerned, the course followed by the participants relies heavily on the expression of opinions. The phrase *j’aime* (‘I like’) is introduced very early and is heard and seen in almost every lesson. The course also centres around teaching students to talk about themselves. As a result students are overexposed to the pronoun *je* (I) at the expense of using the third person (C. Bell et al., 2016).

Ellis, O’Donnell, and Römer (2014) made similar observations in their research of 393 advanced learners of English, correlating frequency of input with frequency of output, in a single experiment. The fact that their participants – advanced learners of English – responded to the input, has implications for my findings. I had hypothesised in section 4.4.3 that participants’ growing vocabulary explained why the frequency of ‘I like’ had diminished; but the paper cited above suggests that a reduced frequency of input in the classroom (which I did not measure, but seems reasonable given the curriculum) might have been a factor. This example of transfer of training is a powerful illustration of the usage-based principle of emergentism, in which the interaction of two simple elements (here, the over-exposure to both the first person, and to opinions) gives rise to a new phenomenon (here, a tendency to translate anything unknown as ‘I like’) (Ellis & Larsen-Freeman, 2006), and is indicative of the existence of a broader complex dynamic system

at play – one in which the curriculum interferes with the participants’ interpretation of the input.

Viewed from a CDST perspective, the initial overapplication of ‘I like’ could also be seen as an attractor state – that is, a patterned outcome which has emerged as many variables (here, different verbs) have spontaneously coalesced (Larsen-Freeman & Cameron, 2008), and with time, there is evidence that the system is beginning to reorganise to allow for new patterns of behaviour (Hiver & Al-Hoorie, 2020) .

10.2.3. Relationship between listening proficiency and grammar

It is harder to draw conclusions as to the relationships between grammatical knowledge and listening proficiency, as research has in some cases been contradictory. Andringa, Olsthoorn, Van Beuningen, Schoonen, and Hulstijn (2012), Mecartty (2000), and Vafae and Suzuki (2020) all detected a relationship, yet Hui and Godfroid (2020) found no robust contribution of parsing skills to L2 listening, although through use of a different instrument. However, my whole-cohort data demonstrated a weak correlation between listening proficiency and grammatical knowledge throughout much of my own data collection period.

In Chapter 4, answering Research Question 1, the perception of content and function words were discussed as a principle grammatical error, followed by difficulty recognising morphological features. To summarise:

- Where a participant encountered a difficult word, content words are more likely to be wrong, yet function words are more likely to be missing, although there was some evidence that participants become better at recognising function words over time.

- World knowledge, general knowledge or common sense seemed to be a greater driver of participants' choice of morpheme than input (this will be discussed at more length in section 10.2.4).
- There is no evidence in the dataset of progress in correct recognition of verb endings to indicate tense.

Both Field (2008a) and Ellis (2017) remind us of the hypothesis that functors *should* be more recognised due to their high frequency, yet also found that listeners handle them differently to content words: in English function words are brief and lack perceptual prominence. If the L2 listener is looking for strong, stressed syllables, they will struggle. Patterson (2021, 2022) suggests that the perceptual problems might be accounted for simply by word length, rather than function, a quality that has not been interrogated further in the present study. But it could be that this perceptual prominence argument depends on the prosody of the language. With a stress-timed language such as English, the unstressed function words are more hidden than in a syllable-timed language such as French, where, in theory at least, their perceptual prominence differs less than that of content words. However, the data from the present study appear to disprove this: function words appear hidden to my participants, too, and this is consistent with the findings of Graham and Santos (2013), who also found English learners of French struggling to discern function words.

Accepting, then, that the differing prosody of French and English appears not to influence the perception of content versus function words, I note that Sheppard and Butler (2017), whose participants had L1s of Chinese, Japanese and Arabic, and L2 English, made a similar finding to mine: their participants correctly transcribed 76% of content words but only 54% of function words. This general picture is consistent with my own – content words are more salient than function words – yet their participants

transcribed many more function words than mine, which I hypothesise might have been due to their higher proficiency. Findings were consistent in a similar study, whose participants were also intermediate and advanced listeners of L1 Chinese and L2 English (Yeldham, 2016).

I was unable to find specific research investigating increased perception of function words over time, but the papers cited above, compared with my own findings, suggest that with time and proficiency, more function words are detected.

If the content / function literature suggests that complete function words can be difficult for listeners to detect, it is reasonable to hypothesise that grammatical morphemes are going to present a similar challenge in a more extreme form (N. Ellis, 2017). As discussed in section 2.4.3 of the Literature Review, there is minimal work on the relationship between grammatical knowledge and listening proficiency, with the exception of that of Mecartty (2000), who failed to find a significant relationship between the two variables. Rost (2015:38) sets out a potential model of how a listener parses incoming language, suggesting a first parse for content and a second for grammar, in which the listener ‘assigns all words into grammatical categories, assigns structural and semantic relations between them’. In this passage, Rost is assuming that the listener can identify words correctly; I have demonstrated that my participants do not do this with any consistency, and as such it follows that Rost’s ‘second parse’ would be an impossibility for them. Instead we see piecemeal and erratic attempts at parsing, and the ever-present influence of other factors which contribute to the translations and mistranslations that are made. This would go some way to explaining why the ongoing relationship between grammar and listening proficiency is markedly weaker than that of vocabulary and listening, and implies a potential perspective of grammatical knowledge

being contingent on vocabulary knowledge: a finding consistent with those of Lowie and Verspoor (2015) and Van Geert (1991).

10.2.4. Metacognitive systems at group level

In Findings 2 section 5.3.1. I showed that all four elements of the MALQ had mean responses that were remarkably consistent, sitting between ‘agree a bit’ and ‘agree quite strongly’, with only very small fluctuations in bandwidth. I also showed that at group level the four elements of the MALQ could be split into two distinct pairs: Problem-Solving and Mental Translation, and Directed Attention and Person Knowledge. Finally, I found that moving correlations between listening proficiency and the elements of the MALQ remained fairly steady throughout the study period: moving correlations (copied from in section 5.3.2) are given in Figure 115, and for ease of reading, separated out into individual graphs in Figure 116. It demonstrates that the correlations between listening proficiency and both Mental Translation and Problem-Solving are almost always moderate, whereas those between listening proficiency and both Person Knowledge and Directed Attention are almost always weak.

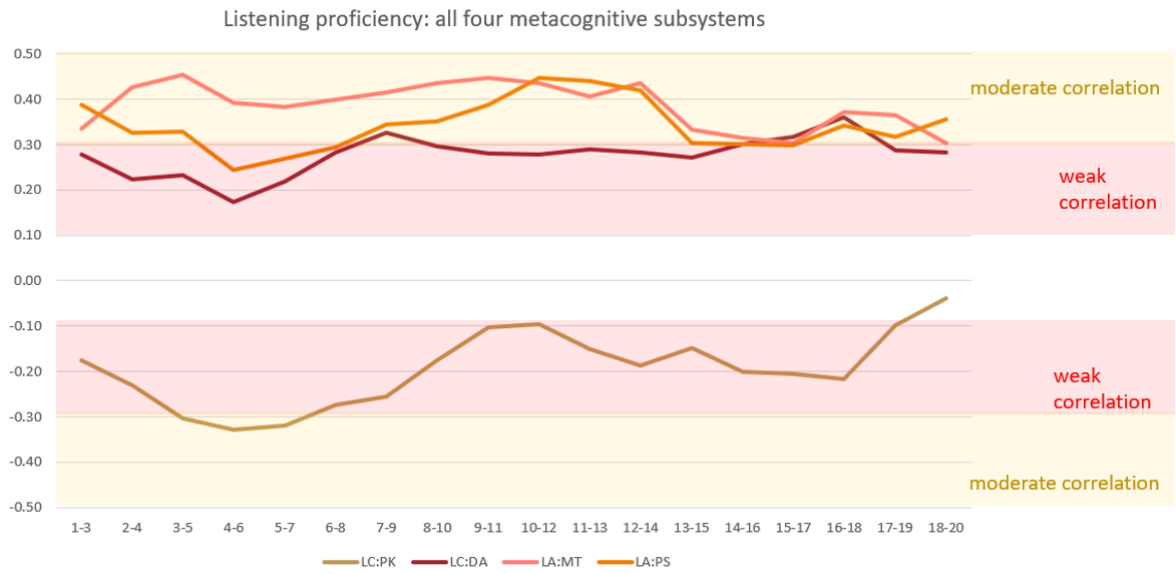


Figure 115: All four metacognitive subsystems and their moving correlations with listening proficiency

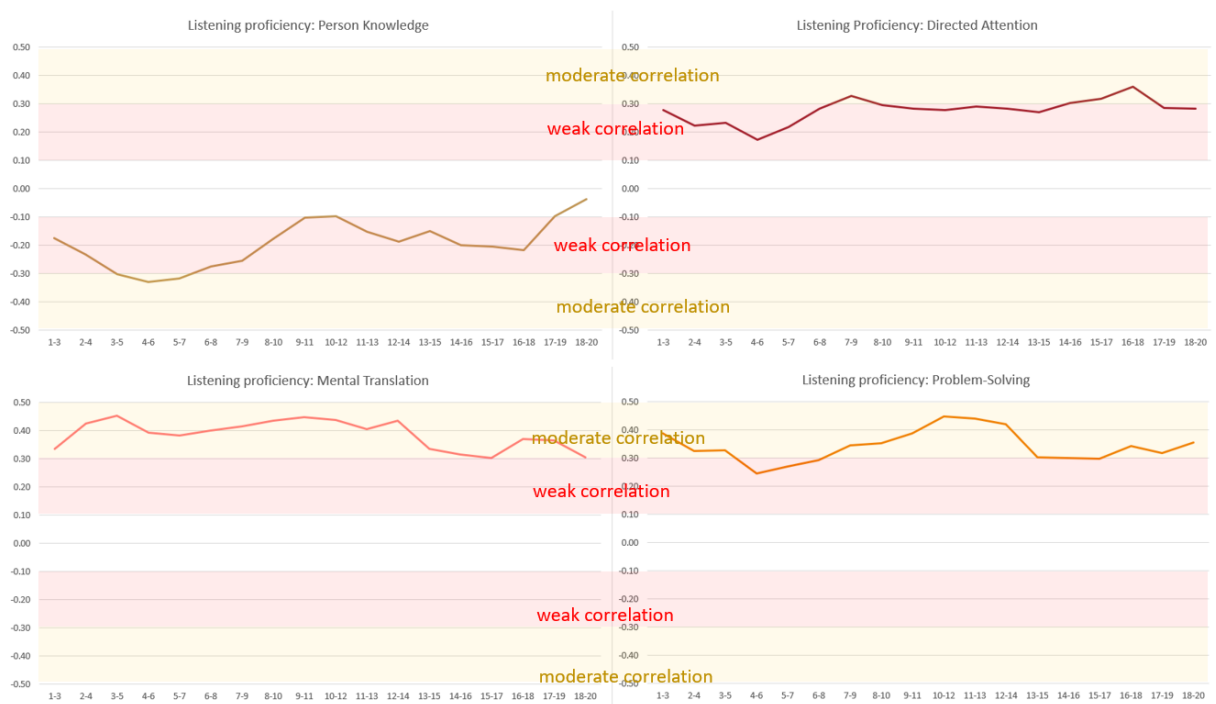


Figure 116: Recap of moving correlations for whole-cohort MALQ

Section 2.4.4 from the Literature Review demonstrated the wealth of literature on the role of metacognition in listening, including a body of work which addressed metacognition longitudinally (set out in Table 1).

Given the inevitability of gaps in L2 listeners' knowledge and understanding of the language, they need to use strategies to compensate (Graham, Santos and Vanderplank, 2010), and the findings recapped above are consistent with this assertion. It is also relevant to recall the work of Vandergrift and Baker (2015, 2018), who found that among 13-year-olds at the end of their first academic year of French immersion in Canada, metacognition was a strong predictor of listening success, but for 10-year-olds only two months into their first year of French immersion found the same link to be weak and non-significant. My own participants are close to the thirteen-year-olds in terms of age, but their proficiency is closer to that of their 10-year-olds; and indeed the findings illustrated in Figure 115 are consistent with theirs in that my links are also weak. This might suggest that development is more about proficiency than age.

10.3. Case studies over time

I reported a wealth of findings from my case studies in chapters 6 to 9. In this section I will discuss what I consider to be the key findings, or ones which are of particular interest. When considering the learning experiences of Helen, Theo, Fran and Ryan, I will make no attempt to separate out this section into the two sets of subsystems as I have done above for the group-level discussion. For example, when considering vocabulary knowledge in L2 listening through the lens of individual differences, van Zeeland and Schmidt (2013) found a range of levels of contribution of L1 vocabulary knowledge, which might be related to different types of compensatory strategies (Stæhr, 2009; Bonk, 2000; van Zeeland, 2014). Cao et al. (2016) identified a 'snowball effect' whereby one listening difficulty triggers others, and for the present study this can be explored through the methodological choice of regular semi-structured interviews and the stimulated recall

that took place at each time point. The interplay of the metacognitive and linguistic systems becomes highly apparent and the CDST approach comes into its own.

10.3.1. The ergodicity problem

When considering each of the case studies, it is striking how little their experiences resemble the group averages discussed above: this is apparent when comparing the group-level rolling correlations with those of the individuals. This is known as the ergodicity problem (Ducker, 2022; Lowie & Verspoor, 2018), whereby group level results do not paint an accurate picture of any one individual and in the case of my research can be seen in graphical form in Figure 117, (which duplicates Figure 51: Whole cohort moving correlation between listening and vocabulary, and listening and grammar Figure 118 (which duplicates Figure 65, Figure 80, Figure 94, Figure 108), Figure 119 (duplicating Figure 56), and Figure 120 (duplicating Figure 70, Figure 84, Figure 98, Figure 112). Lowie and Verspoor (2018), researching the L2 English writing progress of a group of 22 Dutch teenagers over a period of a year, were not even able to form the participants into ‘ergodic ensembles’ – that is to say, into smaller groups whose progress resembled each other. Similar findings have been made by other researchers in the CDST field such as Wirtz and Pfenninger (2023) and is something of a recurring theme in CDST methodological literature. In my own research, the individual developmental trajectories are so dramatically different from that of the whole cohort that I am not even able to draw conclusions as to phases of development, which I showed at group level in sections 5.2.3 and 5.3.2.

It is also relevant to note that the finding from the whole cohort – consistent with much previous literature – that vocabulary knowledge largely correlates with listening proficiency, is not at all visible through the individuals’ findings shown in Figure 118.

The methodological implications of the ergodicity problem will be explored in Implications Chapter section 11.2.4.

I have attempted to interpret these data in the relevant findings chapters. However, at this juncture the charts' presence is simply to illustrate the degree to which the full cohort data flatten out the extremes of the individual experience, and by extension, therefore, give value to the importance of the case studies, as it suggests that we cannot necessarily draw conclusions that apply to individuals through use of the whole-cohort data. Where researchers often take a precise and careful positivist approach to whole-cohort research, the ergodicity problem suggests, to paraphrase my frontispiece precision might not always equate to truth on an individual level.

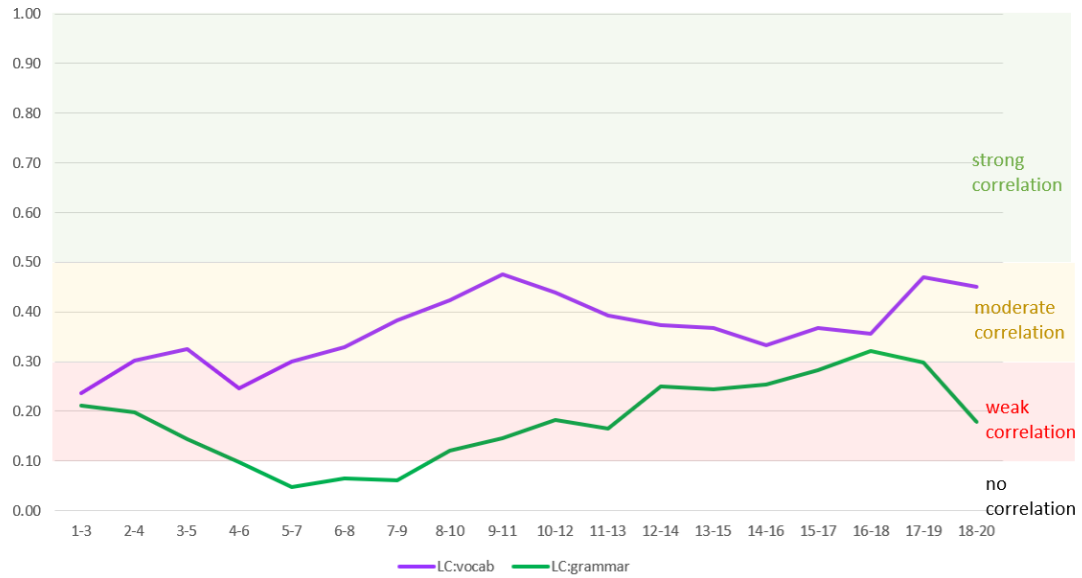


Figure 117: Moving correlations between the linguistic subsystems at group level (to be compared with that of the individuals)

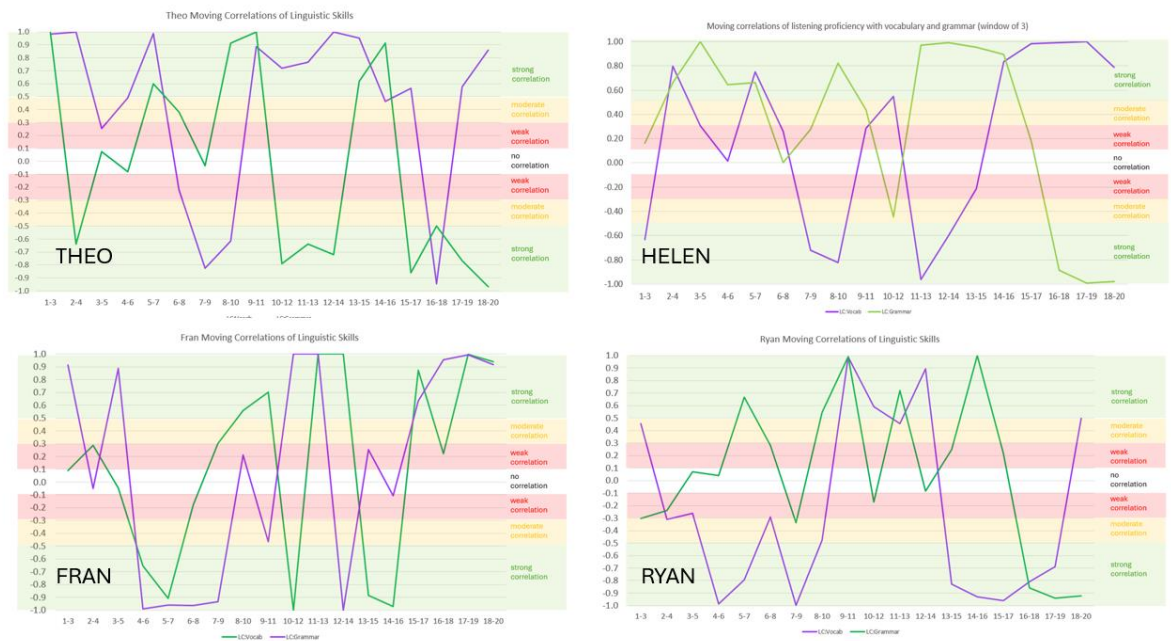


Figure 118: Moving correlations between listening proficiency and vocabulary and grammar: case studies

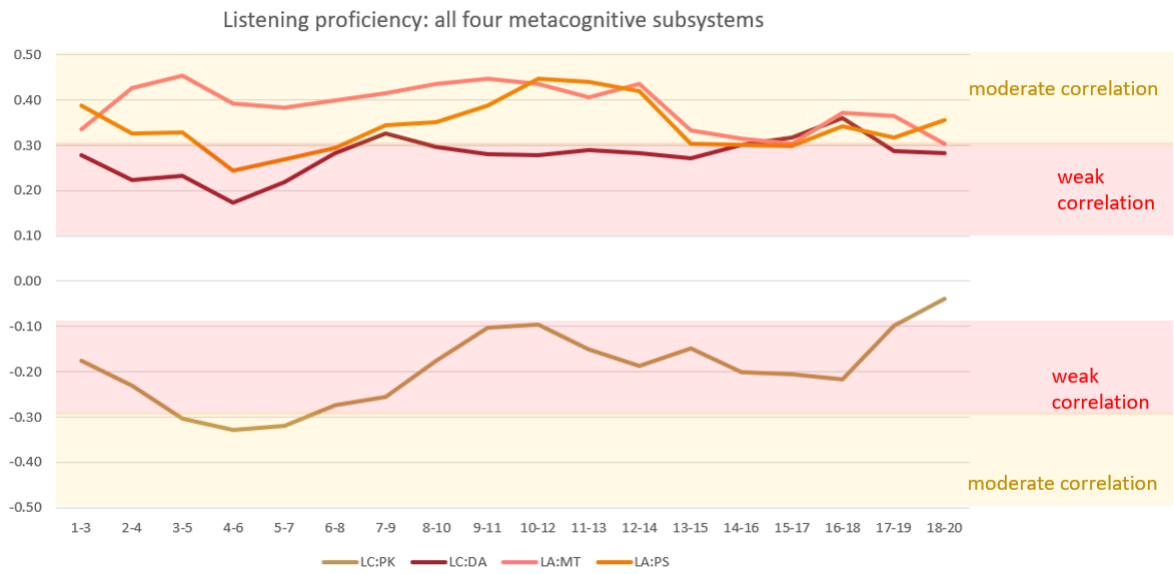


Figure 119: Whole cohort moving correlations for the metacognitive subsystems (for comparison with the case studies)

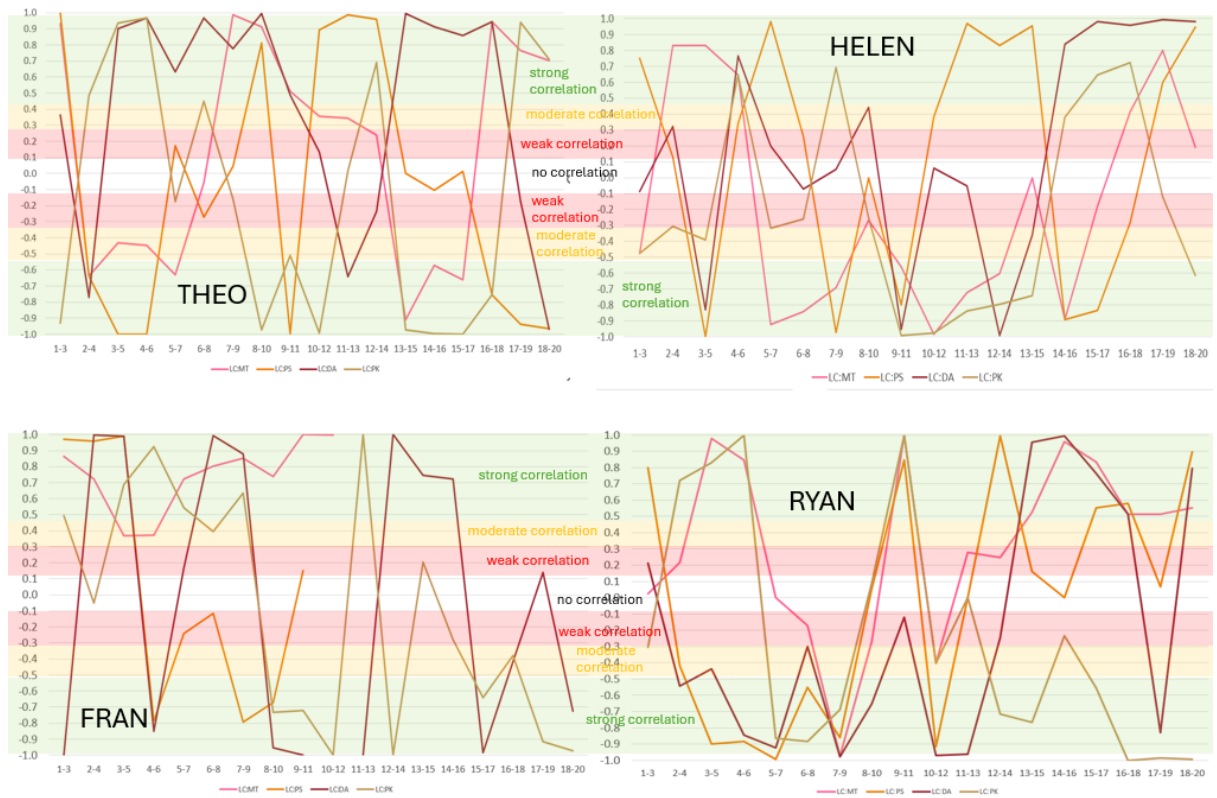


Figure 120: Case studies moving correlations: metacognitive subsystems

10.3.2. Helen as ‘self-subtitler’ (reprise)

In chapter 6 I presented the finding that Helen’s approach to listening was distinctive in terms of her inclination towards mentally representing words that she heard.

Helen’s mental lexical representation (what I called ‘self-subtitling’ or ‘karaoke listening’ in section 6.7.2) bears some resemblance to the work of Heavey and Hurlburt (2008).

Their paper on The Phenomena of Inner Experience explored the mental representations of daily life of thirty participants and within the wide variety of different perspectives, that some reported a sub-category of ‘inner seeing’. Although they did not ask their participants to report whether they saw words in their mind’s eye when they heard them, the authors touch upon the phenomenon and within their sample it is not reported as particularly unusual⁷⁷.

However, the research of Heavey and Hurlburt (2008) does not concern itself with language use *per se*: therefore it is significant to cross-reference the construct of ‘inner seeing’ with two other fields of research: firstly, with Perfetti’s Lexical Quality Hypothesis (Perfetti, 2017; Perfetti & Hart, 2002), which centres the importance for comprehension of both orthographical and phonological specifications of a word. And secondly one might also call on the work of Vanderplank (e.g. Vanderplank, 2016), as well as the meta-analysis by Montero Perez, Van Den Noortgate, and Desmet (2013) who have shown that use of same-language subtitles or captions enhanced participants’ online understanding as well as longer term language skills and cultural understanding.

⁷⁷ Anecdotally, having discovered this phenomenon through regular discussion with Helen, I began to ask other classes that I teach whether they did the same thing. In most classes at least one student experienced this phenomenon, to varying degrees of intensity.

Helen's mental representation of the lexicon might bear some resemblance, in her mind's eye, to the captioned listening experience discussed above. This speaks to an inherent entanglement of the two skills of listening and reading and a possibility that for different learners the skills interact in different manners, but it also tests key theories of language learning. Work on captioning was initially inspired by Krashen's Comprehensible Input Hypothesis (Krashen, 1985), the hypothesis being that input would be made more comprehensible if the learner were permitted to access both modalities simultaneously. It was subsequently developed through the Noticing Hypothesis (Schmidt, 1990, 2001, 2010): where captions complement auditory input, it is more likely to become intake, through conscious attention. Helen did not claim to be able to make a conscious decision to self-subtitle what she heard, and there is no way of knowing the accuracy of her own internal lexical representations. However, the mental phenomenon she experienced did appear to focus some conscious attention on the input and therefore might have had some impact on her own developmental trajectory of input to intake.

Helen's own propensity to 'see' the words she was hearing in her mind's eye changed as her understanding of French developed, however, becoming a less prominent aspect of her self-reported listening over time, then more significant again during the covid lockdowns. In our semi-structured interviews I would ask Helen what was happening in her mind as she listened, but I tried to avoid yes or no questions, which meant that I did not expressly ask her whether she was still 'seeing' words. This means I do not have precise data on when this phenomenon began and ended for her, meaning it is not possible for me to map this reliably onto the developmental trajectory of her vocabulary knowledge or onto the rolling correlation between listening proficiency and vocabulary knowledge. However, it appears that it might be a precursor to a growing (and ebbing and flowing) automaticity of understanding.

10.3.3. Helen and Theo: Alternating between vocabulary and grammar

I addressed the discrepancies between the case studies' data and the full cohort in the Ergodicity section above (section 10.3.1). However, the fluctuations in scores particularly in the cases of Helen and Theo are worthy of further discussion. In chapter 6 I presented the findings that Helen's developmental trajectory of linguistic subsystems appeared to switch between favouring use of vocabulary or grammar, but that rarely both correlated positively with her listening proficiency, and I made a similar finding with regard to Theo which is set out in section 7.3.3 of chapter 7. Furthermore, I noted that both participants even appeared to show some regression in her grammatical knowledge towards the end of the study.

Indeed, such a variability in vocabulary knowledge and frequently opposite variability in grammatical knowledge was visible both in the quantitative and qualitative data, for example Helen's successful translation of *j'aime regarder* as 'I like to watch' at Time 9, reverting to 'I'm going to watch' at Times 13 and 18, and a translation at Time 11 of *nous* ('we') as 'I', with a similar erroneous default to the first person at Time 15 (examples in section 6.4.5 in Findings 6).

When asking the question why might my two high-prior attainers, whose listening proficiency was consistently above average, show a regression of grammatical knowledge, it is helpful to refer back to complex dynamic systems theory. For example, Lowie and Verspoor (2018) demonstrated that their participants with the highest scores for aptitude, motivation and exposure (when studying the developmental trajectory of L2 writing) showed more variability than the less proficient participants, and this finding is consistent with other CDST work in a similar vein (e.g. Verspoor & Bot, 2022; Yu &

Lowie, 2020), as variability is the key component of change, and therefore, by extension, of learning, or progress.

From the CDST perspective one might also argue that when the variability shows a progress upwards followed by a regression, the variability is indicative of knowledge which is being acquired but is not yet secure enough to be *consistently* demonstrated. In the cases of Helen and Theo, this might be due to the cognitive load that is being negotiated at the time, where sometimes lexical retrieval might dominate, at the expense of successful parsing.

10.3.4. Theo and Fran: Home languages, affect, and life capital

Life capital, a research perspective coined by Consoli (2021, 2022) is a person's symbolic 'wealth' brought about by their life experiences or life story. The author acknowledges the value of CDST to 'offer a complex and holistic perspective' (Consoli 2022:4) but claims that it risks overlooking 'the humanity and unique lives of the participants'. In this section, I discuss the additional impact of both Fran's and Theo's home lives and affects (which, indeed, I demonstrated to be connected in chapters 7 and 8). I argue that by taking a retrospective case study approach to the complex dynamic systems of my participants' listening trajectory I can demonstrate the elegance with which the construct of life capital can be integrated into CDST research.

Fran and Theo were both exposed to additional languages at home. Fran's fluent Italian frequently impacted linguistically on both her understanding, and her development of French and I showed in section 8.4.2 how it was initially a hindrance and later a help. In a similar vein, Theo attributed his strategic use of verbs in understanding French in part to his mother's use of Swahili verbs when asking him to help at home. Both also alluded to a relationship they had between their motivation to succeed and a positive relationship

with their families – Theo briefly and Fran on a regular basis, and it is clear that for both, then, affect formed a key part of the system which was instrumental in their progress.

Indeed I discussed in section 8.7 how, when discussing strategy use with Fran at interview, conversation often turned to questions of self-regulation: her approach of taking a deep breath before listening was also cited by the participants of Cross (2009).

10.3.5. Remote learning and its implications

In terms of the whole cohort, the timings of the first covid lockdowns did not coincide with a drop in my recorded listening proficiency, vocabulary knowledge or grammatical knowledge, although the second lockdown did coincide with the beginning of a dropping off in vocabulary knowledge. No relationship was apparent between the lockdowns and the whole cohort's metacognitive subsystems.

From a case study perspective, though, while Helen seemed to take the lockdowns in her stride, Theo, Fran and Ryan all faced challenges. Theo's appeared to be purely pedagogical: he demonstrably regressed in his listening skills and discussed this at interview, addressing how he felt that my listening surveys were harder as he was out of practice with listening to French. (For practical reasons it had been very difficult to give the students listening activities during lockdown work.)

By contrast, Fran struggled emotionally during the first lockdown, which impacted on her learning. Her self-imposed pressure to keep up with work went head-to-head with a growing lethargy, and eventually it was the lethargy that triumphed: we can see this in her lack of data at Time 12. Ryan struggled with his organisational skills, which resulted in minimal work being completed at home, although he did not report feeling stressed or upset by either the lack of work or indeed the situation of the pandemic more broadly – a finding that is consistent with those of Güvercin, Kesici, and Akbaşlı (2022), who

reported that distance education did not cause angst in some cases due to indifferent parents or special educational needs. I will discuss Ryan's case in more detail in section 10.3.6, below, given the interaction of covid and his special educational needs.

It is worth discussing the question of remote learning despite the fact that it does not address any of my research questions (which were crafted before the pandemic). Still, the present study is remarkable in that it – inadvertently – collected data on learning throughout the pandemic and as such there are important findings which might be pedagogically pertinent in other remote learning situations. Aside from future global pandemics, this could include:

- Geo-political barriers to attending school, such as war or major infrastructure disasters.
- Emotional or medical barriers to attending school, such as long-term health problems or emotionally-based school avoidance.
- Cultural barriers to attending school, such as children from travelling communities.
- Geographical barriers to attending physical school, perhaps due to extreme distance, combined with a sense that online school is preferable to boarding.

The annual Language Trends report from 2021 (Collen, 2021) reported that two in five pupils of the ages of my participants struggled to engage in online learning, which seems to be slightly higher than was the case for my own study. Recall as stated in Methodology section 3.8.2 that I disappplied a number of participants who simply stopped responding to my listening surveys when they had to complete them at home. Collen's report highlighted the particular struggles that students had in accessing speaking and listening tasks during this time, and also addresses the inevitable reduction in trips abroad. Indeed,

it cites 88 per cent of schools reporting a negative impact on language learning during the first national lockdown: 62 per cent reporting a ‘big negative impact’ and a further 26 per cent saying there was a ‘small negative impact’ (p.4). This perception was not confined to the UK: teachers across the globe reported specific difficulties with online language learning, including slow progress due to the lack of face-to-face teaching (Klimova, 2021) and worries about the use of translation software (Mavridi, 2022). Consistent with my own experience, Cattán et al. (2021) found that remote learning improved in the second lockdown (January to March 2021, which coincided with my Time 17, during which online lessons took place for the first time), but it is noteworthy that despite this, my participants’ vocabulary knowledge appeared to fall, as recorded by the AuralLex (see section 5.2.1 for data).

A key implication for addressing future remote learning eventualities, then, is the need for sensitivity to individual circumstances. Not all students will react in the same way to school closures and family circumstances, including practicalities such as access to the internet or computers will vary widely. Makhovych (2024), in the context of online L2 English learning during the war in Ukraine, discusses addressing these issues through the gamification possible through the internet as well as the importance of allowing flexibility, where not all students can access work simultaneously. And, given the risk of post-traumatic stress after pandemic-related quarantine (Sprang & Silman, 2013), the key implication of remote learning must be to value the individual and address needs on a case-by-case basis.

10.3.6. Ryan: Special needs, covid, and a widening ‘gap’

In Ryan’s case, I found a student who did not appear to make progress in his French listening over the three years of the study, as far as the quantitative data were concerned,

although there was evidence of progress through our regular interviews. Ryan was diagnosed with a special educational need during the period of the study, and found it very difficult to complete work during the covid lockdowns so much so that he was among a very small group of students who were invited back into school by May 2020. This section, then, will discuss special educational needs (SEN) both within the context of language-learning and explore how students with SEN might have been additionally impacted by Covid. I had not planned to have a case study with SEN. Nonetheless, Ryan's inclusion in the work is important, even if this had not been anticipated at the start of the project. It is estimated that during the time of data collection, 15.5% of students in the England were diagnosed with a special educational need (Department for Education, 2020).

The 'Achievement Gap' refers to the difference in educational achievement between various disadvantaged groups and the wider populations and has been evidenced widely (Education Endowment Foundation, 2018, 2022). The widening standard deviation in the listening scores I showed in section 5.1 suggested that an achievement gap developed in my own dataset, and Ryan's own lack of progress was symptomatic of this. The Education Endowment Foundation (2018) measured students reaching the expected standard in reading, writing and mathematics and showed that the widest gap was between those students with SEN (14%) compared with those without (62%). Furthermore, the area in which the present study took place was found to have the widest attainment gap in the country (Education Endowment Foundation, 2018).

One might very reasonably argue that progress in French listening is not an adequate proxy for broader educational progress; in which case, it is noteworthy that my data in the case of Ryan appears to be consistent with much wider nationwide findings.

Despite the potential for a difficult learning experience, some scholars are positive about language learning for students with SEN. Beginning with a subject new to all class members has the potential to ‘level the playing field’ (Howard, 2023; Joy & Murphy, 2012), particularly when students can reset their self-esteem and motivation, and when learning might be focussed on peer interaction and oral communication rather than written work. Additionally, teaching approaches which might consist of ample use of gesture, and plenty of group work and modelling, have been shown to enhance learning for students with additional needs (Mady & Muhling, 2017). It was evident that Ryan’s self-esteem and self-efficacy never appeared to be negatively affected during his learning of French, yet his own experience might also be more consistent with the findings of Andrews, Robinson, and Hutchinson (2017), who reported that SEN students struggle to keep pace with their peers.

The pandemic probably played a part in Ryan’s lack of progress, too. I showed in Figure 100 how Ryan’s listening proficiency had plateaued around the time of the Covid lockdowns, and did not appear to make any more progress after this. The same was broadly true of his AuraGram scores (Figure 105) although it is difficult to draw conclusions from his AuralLex scores (Figure 103) which fluctuated so wildly. Post-pandemic meta-analysis reached a clear conclusion that the lockdowns resulted in an increased attainment gap and that students did not make the progress that they otherwise might have done (Donnelly & Patrinos, 2022, Education Endowment Foundation, 2022, Engzell, Frey, & Verhagen, 2021), with students in Key Stage Three (the key stage of my participants) particularly suffering (Department for Education, 2021, 2022b).

As far as Ryan’s progress was concerned, then, his experience was largely consistent with wider research on cases and situations such as his: students with special educational needs might enjoy language learning but do not necessarily make good progress, and that

the intersection of SEN with the special situation of the pandemic meant that his lack of progress was perhaps no surprise. Implications of this will be discussed below in section 11.5.3.

10.4. Discussion: Closing remarks

In this chapter I have attempted to discuss and analyse some of the most interesting findings generated by the study. I have shown that my findings with regard to listening proficiency and vocabulary knowledge appear to be consistent with the existing literature and probed some of the reasons for this through the error analysis. I have also shown how the development of errors in auditory grammatical knowledge might go some way to explaining the reasons why existing literature is inconclusive about the relationship between grammar and listening proficiency, and have highlighted the value of my longitudinal approach.

I have also shown a key conundrum of CDST research: that of the ergodicity problem, whereby individuals' experiences do not align with that of the group average. And with regards to the case studies, I have addressed the unusual case of Helen's lexical mind's eye, the highly fluctuating dynamism of Helen's and Theo's linguistic subsystems, the role of life capital with regard to learners with additional L1s, special educational needs, and the impact of remote learning brought about by my collecting data throughout the pandemic.

The findings have various concrete implications for research and practice, which will be addressed in the next chapter.

11. Implications

11.1. Introduction

This chapter will address the implications of the present study. Given that an underpinning tenet of CDST research is that it will generalise to theory rather than to populations, I will begin with theoretical implications. Generalisation to theory results in academic work which is hypothesis-forming rather than hypothesis-testing, which means the theoretical implications will inform potential directions for future research.

In addition, my experience of carrying out the research will result in my addressing some methodological and positional implications, and finally I will address pedagogical implications for classroom practice and for resource creation.

11.2. Theoretical implications

I suggested in my literature review that Field's model of listening (2008b, 2013) could be argued to be a complex dynamic system, with a range of subsystems which are constantly interacting. The balance between the system changes according to the constantly evolving input, and what a listener understands is an iterative, dynamic process dependent on a variety of external factors, such as context and environment. However, where much of the current literature in the field of L2 listening aims to examine the relationship between one subsystem and 'the whole' (usually seen as success in listening comprehension, operationalised by comprehension questions or by free recall), this CDST approach to listening takes a more holistic view in order to understand more about the constantly changing relationship between the subsystems as well as broader factors such as the listener's self-efficacy and environment (De Bot et al., 2007).

With my examples across the dataset and within the context of various subsystems, I have endeavoured to present a theoretical perspective whereby a complex dynamic system of second language listening can play out over three seconds or three years, and that each participant's experience of listening to each individual track is a complex dynamic system played out on a micro scale whereby a trajectory of understanding develops over a matter of seconds.

All the expected characteristics of a complex dynamic system, as discussed in the previous literature, are apparent. The findings above clearly demonstrate the importance of the relationships between different sounds and words in each track, and how the presence of one influences the interpretation of others. We see emergentism: the 'spontaneous occurrence of something new' (van Geert, 2008:182) such as 'skinning people' for *ski nautique*, where components of the system interact and self-organise 'without direction from external factors and without a plan of the order embedded in any individual component' (Mitchell, 2003:6) – for example in the case of *seulement / Solomon* (see section 4.5.1). We see non-linearity, in that a small change in one parameter can have significant downstream implications. Take the introductory example from Chapter 4, which answered Research Question 1, where one participant translated

sa sœur Martine est très travailleuse

/sa sœʁ mɑʁtin_ε tʁɛ tʁavajøz/

(‘his sister Martine is very hard-working’),

as ‘Martin has nice eyes.’ The salience of the misunderstood name ‘Martin’ had implications in that it forced the participant to ignore the input of *sa sœur* (indicating female gender) and instead followed the logic of their own translation. At the data analysis stage of a CDST study, ‘noise’ (defined as non-linear progress or unexpected

outliers) in the data is not necessarily interpreted as measurement error, but instead as ‘sound’ (Marin & Peltzer-Karpf, 2009), to be observed and interpreted where appropriate. This is due to the fact that CDST is a theory of process rather than product (Caspi & Lowie, 2013): if one accepts that underlying development is the interaction of a range of subsystems, it could be that what might otherwise be labelled as ‘noise’ in the data in fact gives us information about the system, and I will illustrate this regularly in this chapter.

11.2.1. Field’s listening model and its subsystems

The key theoretical framework underpinning the present study is that of Field’s model of listening (2008b, 2013), which I set out initially in Figure 1 in chapter 1 and have repeated in Figure 121 for ease of reference. I also discussed Anderson’s (1990) three-phase model, which is made up of ‘perceptual processing’ (i.e. hearing a sound as a familiar language), ‘parsing’ (in which the listener breaks the sounds into words and chunks), and ‘utilisation’ (i.e. giving meaning to words and chunks), and that of Cutler and Clifton (1999) which divides the same process into four operations: ‘decoding’, ‘segmenting’, ‘recognising’ and ‘utilising’ where Anderson’s ‘parsing’ becomes the two elements ‘segmenting’ and ‘recognising’. Where Anderson refers to ‘phases’ and Cutler and Clifton to ‘operations’, Field (2013:94) prefers ‘levels of analysis, which better reflects the recursive nature of listening’, the iterative process of which I have demonstrated in all the findings chapters.

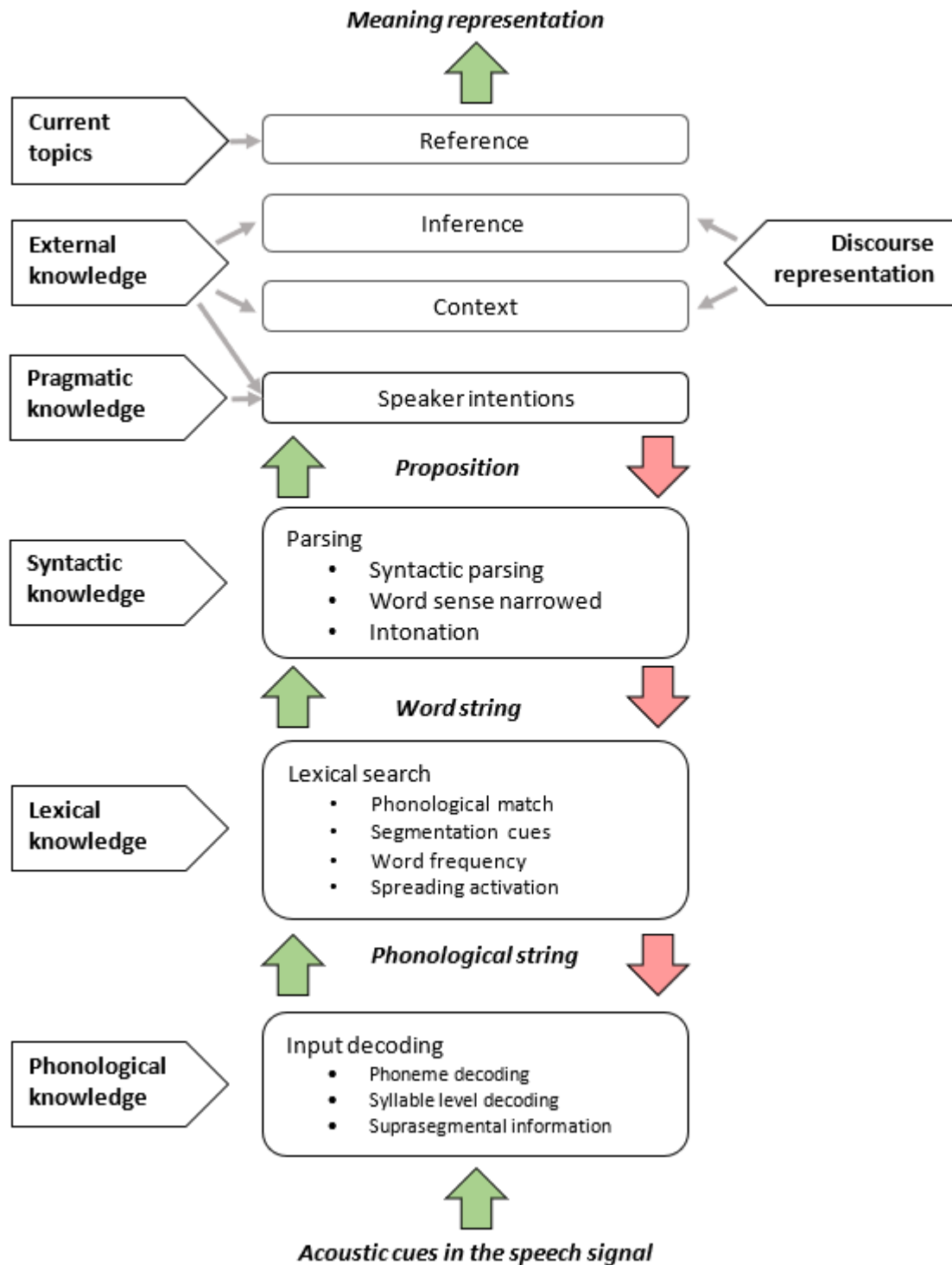


Figure 121: Field's model of listening (2008b, 2013:97; 2013:101)

I wrote a first draft of my literature review in the first half of 2018 and began data collection in September of that year. From the point of view of the interplay between phonological and lexical knowledge in Field's model, in 2023, Saito, Uchihara,

Takizawa, and Suzukida, published a paper developing Nation's (2013) perspective on knowing a spoken word which consisted of 'phonologisation' (recognition of phonological forms without orthographic cues), 'generalisation' (ability to retrieve lexical knowledge) and 'automatisation' (accessing meaning and collocational aspects of the word rapidly and stably), and imply that L2 listening is characterised by different levels of each phase, with the least proficient learners favouring the phonologisation phase.

I demonstrated in all my findings chapters that the relationships between listening proficiency and grammatical knowledge was inconsistent over time, and throughout the three years very little parsing was apparent. Adding this fact to the recent work by Saito et al. (2023), one key theoretical implication is that while Field's model remains appropriate, the relative weight of the various subsystems varies according to individuals and proficiency. The model illustrated in Figure 1/Figure 121 also contains arrows. The specific order of the levels of analysis suggests that although they are iterative, the listener is obliged to pass through them in order. I would now like to suggest a revision of Field's model for my beginner participants, which is illustrated in Figure 122. This revision brings with it scholarly implications for further research which will be explored in section 11.3.

11.2.2. Listening to learn: Noticing grammar

At the beginning of the Literature Review I discussed the difference between 'learning to listen' and 'listening to learn'. While the emphasis in the present study concerns itself with the developmental trajectory of listening proficiency – in other words, learning to listen – it is also important to consider that one purpose of listening to an L2 is in order to enhance one's proficiency in the language – in other words, 'listening to learn'.

I demonstrated throughout the findings and discussion a relationship between vocabulary knowledge and listening proficiency, but erratic at best between grammatical knowledge and listening proficiency. There is a key theoretical implication here in terms of testing Field's model of listening (2008b, 2013), which is the key theoretical framework underpinning the present study.

There is also a theoretical implication with regard to the Noticing Hypothesis (Schmidt, 1990, 2001, 2010), which states that learners must be able to notice and attend to input if they are to be able to output it. My findings suggest that for my beginner learners, at least, the Noticing Hypothesis might either not apply at all, or at best depend on input in a written modality, the implication being that there might be a minimum level of proficiency – or other factors – necessary for different types of Noticing to occur. Given these dual theoretical implications, I propose an adaptation of Field's model, which is set out in Figure 122.

Within my adaptation of Field's model of listening for my beginner to early intermediate young learners of French, different parts of Field's stages of analysis carry different weightings. Phonological Knowledge dominates, with Context and Current Topics also playing a role. Lexical Knowledge features but is less important; the roles of Syntactic Knowledge, Pragmatic Knowledge and Discourse Representation are minimal. While Field's model contains arrows moving in both directions, I saw no evidence that any intake gained from Parsing would feed back into enhanced Lexical Knowledge. Instead, a cycle between Context and Lexical Knowledge arose, largely bypassing Parsing. It is significant that Hui and Godfroid (2020) also report that their participants appeared to bypass syntactic parsing. Even Field himself acknowledges that not all input becomes intake (Field, 2008a). Participants were intermediate level learners of English with a

range of L1s, aged in their late teens and early twenties and for these, inflections and verb forms appeared to present difficulties.

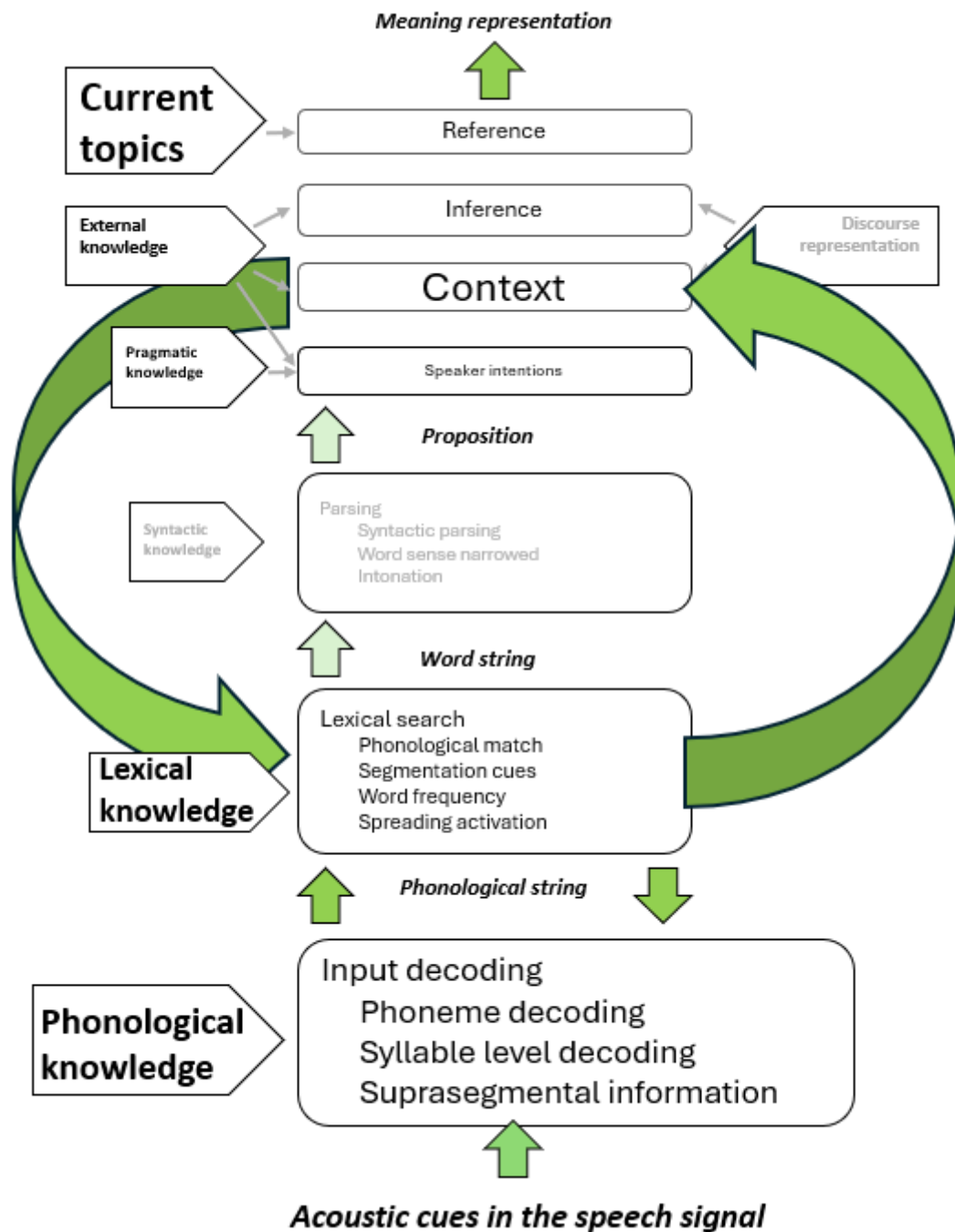


Figure 122: Adaptation of Field's model of listening according to my participants

The theoretical implication, then, is that the different stages of analysis presented by Field in his model of listening (2008b, 2013) carry different weights, and these might be

dictated by the proficiency of the listener, although it is possible that other factors might also be at play.

11.2.3. Receptive interlanguage

Participants' lack of attention to grammatical morphemes in the input implies some importance of modality when applied to the Noticing Hypothesis. In addition, the breadth and depth and myriad idiosyncrasies of the paused translation corpus appears to represent a receptive equivalent of interlanguage (Selinker, 1972). The concepts and underpinnings of interlanguage are inherently linked to those of Complex Dynamic Systems Theory, featuring attractors, emergence, dependence on initial conditions and on context. To combine these ideas, a construct worthy of future investigation, then, might be that of 'receptive interlanguage', the idea whereby each individual's mental representations of what they are understanding (be it in a written or aural modality) evolves according to time and proficiency.

11.2.4. The ergodicity problem (reprise)

The ergodicity problem as discussed in section 10.3.1 of the Discussion chapter carries with it implications, then, on both a theoretical and a pedagogical level, although both have the same underlying question which is – if one accepts that the experience of no single individual matches that of the full group, which type of research is valid?

Lowie and Verspoor (2018) offer a solution by suggesting that the value of the group study is to hint at the relative weight of individual factors influencing second language development, while longitudinal work should be geared towards the case study. As such the two should not be seen as mutually exclusive, but rather complementary (Hiver & Al-hoorie, 2022), with individual case studies generalisable to theory and group studies for hypothesis-testing (Larsen-Freeman & Cameron, 2008). Wirtz and Pfenninger (2023)

have recently developed further complex statistical methods in which they integrate population-level research with both small-group and case studies with the aim of addressing the problem of generalisability: as such it is clear that the field continues to grapple with the ergodicity problem in order to evolve.

11.2.5. Theoretical implications: Mini conclusion

The findings advance our understanding of second language acquisition by challenging notions of ‘static competence’ (Larsen-Freeman, 2017). I have shown that competence in L2 listening is not even static from one track to another, let alone from one time point to another. Instead, participants’ outputs (if indeed that is what we call competence) are in states of constant flux and interaction with its environment. This obliges us to challenge conventional dichotomies (e.g. Field, 2008a), whereby classical experiments simply ask what is the relationship between x and y . Instead we might attempt to account for the non-duality of any phenomena under investigation and become more comfortable with a model of perpetually moving parts, with the theoretical implications that this carries with it. We cannot be sure that precision equals truth.

11.3. Scholarly implications

The collection of future research ideas here speaks to the fact that the present study aims to form rather than test hypotheses. Table 54, below, sets out a range of findings that have arisen from the study and the ensuing hypotheses that could now be tested.

Table 54: Implications for future research

My finding	New hypotheses to be tested
False friends are problematic	There a proficiency level at which false friends cease to be problematic. False friends present a greater problem in listening compared with reading. Distance between languages influences how listeners deal with false friends. Speech recognition software will struggle with false friends.
Lexical sets are hard, but some sets are harder than others	Different kinds of lexical sets are harder than others. Difficulty of lexical sets is influenced by their word class. Lexical sets present a greater problem in listening compared with reading.
Sometimes words are translated twice	Doubling-up errors are more likely to occur in reading than in listening Doubling-up errors in listening occur at an intermediate proficiency level. There is a relationship between listening doubling-up errors and vocabulary size.
Paused translation generates rich data	Paused translation tells us more about the listener's experience than paused transcription.
Grammatical morphemes are rarely noticed	Specific grammatically-focussed pedagogical approaches will optimise listeners' intake from auditory input.
Voiced / unvoiced consonants are easily confused	Beginner and intermediate listeners do not discriminate between voiced and unvoiced consonants. The timing of the voice onset in a listener's L1 does not impact on their perception of voice in the L2. To what extent are both these hypotheses dependent on language pairing?
Listening metathesis can be a problem	Listeners are prone to metathesis.
Students find listening harder than teachers imagine	Teachers overestimate beginner and intermediate listeners' proficiency in listening.
Inner orthographic representation of auditory input	'Self-subtitling' affects a small percentage of listeners. 'Self-subtitling' is more common among listeners of specific orientations (levels of proficiency, interacting with strategy use, other types of 'inner voice' orientation such as aphantasia or hyperphantasia)
Vocabulary size and grammatical proficiency grow together, but diverge from Time 16	Intermediate listening proficiency is the trigger for divergence between grammatical and vocabulary proficiency. Change in metacognition (as tested by the MALQ) is the trigger for divergence between grammatical and vocabulary proficiency.
In the MALQ, Problem-Solving and Mental Translation evolved together, as did Directed Attention and Person Knowledge	Problem-Solving and Mental Translation strategies are supportive of each other. Directed Attention and Person Knowledge strategies are supportive of each other.
[not researched in the current data]	Mental translation strategies correlate with vocabulary knowledge for beginner and intermediate learners but not for advanced learners.
Helen and Theo do not demonstrate the 'sound trumps sense' phenomenon.	Less proficient learners use phonology to build meaning, regardless of context. More proficient learners (even beginner learners) do not.
Fran's Italian is both help and hindrance	A second L1 is helpful in L2 listening. The closer the second L1 is to the L2, the more helpful it is.
Ryan made minimal discernible progress in French listening over three years.	Students with particular challenges can learn in input-poor conditions, but will need specific pedagogical approaches.

11.4. Methodological implications

It is significant that the methods described in chapter 3, while derived from previous practice, are in some cases unique, given the paucity of longitudinal listening research, listening research from a CDST perspective, and listening error analysis. As such, the present study contributes to the body of methods with regard to both L2 listening and CDST, and critically, measuring receptive skills through CDST: many of the methods developed could also be applied to looking at a developmental trajectory of reading. Furthermore, although my chosen timescale was three academic years, the tools could be used in the context of both shorter and longer time scales.

11.4.1. Longitudinal approach

The present study has demonstrated both the value and the inevitable complexity of taking a longitudinal approach, to receptive skills in particular. I have demonstrated how the data collected can fluctuate, sometimes dramatically, between time points, and one conclusion of this might be that cross-sectional research risks being less conclusive than it might be presented as. As a result, it is important to develop methodologies in which longitudinal research can be feasible and practical. More reliable longitudinal methodologies will allow researchers to move away from the assumption that what is measured in the moment might be an indication of development over time.

I have demonstrated that Rasch modelling has the potential to be of use when researching receptive skills over time, and in settings where externally verified tests are unavailable (such as those used by Chang (2017) and Dong (2016)). This might have a particular methodological implication in the field of English Medium Instruction, in which a recurring question is the extent to which the approach is pedagogically valuable from a language-learning perspective (Macaro, Curle, Pun, An, & Dearden, 2018); in this field

there remains a need for a good longitudinal proficiency measure for receptive skills (Zhou & Rose, 2024). This methodological implication therefore connects with my ‘listening to learn’ implication addressed in section 11.2.2.

11.4.2. Ethical implications of being a teacher-researcher

Unfortunately, the present study also demonstrated the difficulty of longitudinal research: as stated in section 3.2.3, the research was made possible by the fact that I was also employed by the school in which the study took place. As a part-time student and part-time teacher, I was in an unusual position to be able to facilitate this. In carrying out the present study I was presented with the ethical implications of the blurring of boundaries between teacher and researcher (Consoli, 2021a) and needed regularly to reflect on the potential risk of any unintentional coercion of students into participants, ensuring that they were free to opt out without repercussions (Galloway, 2016). Being a researcher ‘embedded’ in the research setting undoubtedly provided richer, contextually-informed data but demanded a reflexive approach: future researchers must be aware of the need for constant, honest evaluation of their positionality and the influence of their authority.

If similar long-term research is to be carried out, it is likely that convenience sampling would dictate participants are university students. This, too, brings with it an ethical question: that which asks to what extent are language-learners at university typical of language-learners across the world. Most language learners in the world are not university students (Godfroid & Andringa, 2023), yet the vast proportion of research participants are (Plonsky, 2023) – this would exacerbate the sampling bias problems which beset applied linguistics (Andringa & Godfroid, 2020). Future potential researchers, funders and institutions will need to think creatively about how to address this issue.

11.4.3. Complex Dynamic Systems Theory

The present study shows the potential of looking at the same issue across various scales, and demonstrates that different methodologies might be necessary to do so: it was not possible to scale up the qualitative approach shown in Findings 1 across the three-year time period, nor was it possible to scale down the quantitative approach shown in findings 2 to accommodate a moment-to-moment time period. This is an assertion also made by Ducker (2022), who advocates the value of CDST listening research on the micro time scale, in contrast with previous CDST work in listening such as that of Dong (2016) and Chang and Zhang (2021). While Ducker and I both made use of stimulated recall to attempt to access the micro time scale, our other methodologies diverged: Ducker with specially written software in which the listener could self-report their understanding level while listening, the present study with the paused translation approach. We both showed the fluidity, complexity and dynamism of the process of making meaning from auditory input. Methodologically this shows the potential of the moment-to-moment examination of the listening experience, but also that various approaches are feasible.

11.4.4. Instruments

Paused translation, and Mental Translation

The potential for a short stream of speech to appear to contain an English word or a French-English cognate might be a trigger for lexical or segmentation error. This finding could not have arisen if I had used paused transcription rather than paused translation, and offers important new insights into the experience of L2 beginner listening.

I found that only the Mental Translation metacognitive subsystem showed a consistently moderate correlation with listening proficiency. With regard to participants' propensity to

translate, it was suggested by assessors at a milestone viva that my findings in chapter 4 might be an artefact of the paused translation methodology, and that data generated by the MALQ questions on mental translations were also distorted by the methodology. However, in their paper introducing the MALQ, Vandergrift, Goh, Mareschal, and Tafaghodtari (2006:450) acknowledge that Mental Translation is strategy that ‘beginning-level listeners often feel compelled to use’

In the introduction to findings 1, I discussed research which analysed the nature of various difficulties; and in these previous studies, the dominant methodology is one of paused *transcription* as opposed to translation. But in any L2+ listening experience, all the languages known by the listener play a role, including the L1. There is a notable body of literature on perceptual processing in L2 listening which addresses the exploitation of lexical material in the L1 (e.g. Hall, 2002), the role of interlanguage (e.g. Nation, 2000), the L1 phonological nature of entries in the lexicon (e.g. Wilcox & Medina, 2013) and the consequent L1 interference (e.g. Broersma & Cutler, 2008), leading to unclear phonolexical representations in the L2 (Carney, 2021; Cook et al., 2016; Duyck, 2005). To combine this body of research with that of the listening errors (section 2.3.2) justifies the use of paused translation rather than paused transcription.

There follows an argument for paused *translation*, which allows the researcher to tap into any L1 influence in the L2 listening experience, particularly among beginner listeners ‘compelled to use’ their first language (Vandergrift, Goh, Mareschal, and Tafaghodtari, 2006:450). A key finding of the present study is the ever-present and sometimes unhelpful influence of English (i.e. the L1) when the participants are attempting to make sense of what they hear, and the literature cited above suggests such influence was simply made visible through the paused translation methodology. An interaction – sometimes

unwelcome – between the L1 and the L2 contributes to the fuzzy representation of the input in the listener’s mind, and this can be both at the phoneme and the word level.

To end this section I return to the moving correlation between listening proficiency and Mental Translation which is shown in Figure 116: here we can see a drop in the relationship between the two constructs towards the end of the study. If indeed both Eastman (1991) and Vandergrift et al. (2006) assert that listeners should move away from mental translation as they make progress, the drop in correlation is a positive sign, and one that suggests that the translation strategies used by the participants initially might have had little to do with my paused translation methodology and more to do with an almost instinctive response to hearing a language of which they have only minimal knowledge.

Stimulated recall

The use of stimulated recall in the case study element of the present study provided additional insight: the participants spoke about their own processes and removed a need on my part to guess about these processes based solely on their errors or their test scores. To take this approach adds an additional layer of richness to the research in that it unveils some of their more conscious approaches to the experience of listening. For instance, Helen said to me at Time 14, discussing track 6 of test D31:

‘I couldn’t tell whether he said pars or parents, but I tried to use logic. So I assumed with friends, he’s speaking with his friends. Then I started doubting myself, I thought did he say parole or did he say something I haven’t heard of before?’

The methodological implication, then, of the use of stimulated recall is the power of its triangulation and the degree to which it enhances data collected through other means.

11.4.5. Participants

I addressed the question of sampling bias in section 11.4.1, overlapping as it did in my own case with issues relating to longitudinal research. I reiterate the point here that methodologically the present study has important implications given that the participants were school children, not university students, and that their L1 and not their L2 was English. This puts them in a small minority within the SLA literature (Plonsky, 2023).

Unreliable witnesses, and floor and ceiling effects

I discussed in section 9.1 of Findings 9 that I had been unsure about the inclusion of Ryan within the present study given the repeated unreliability of his accounts. Yet I also stated that I had concluded that this very nature was important. Such participants rarely appear in academic literature but if we do not include them, we risk overlooking a relevant sub-population of language learners. The consequence of overlooking these analytically inconvenient learners is that the academic field will never have the opportunity to understand their own learning experience and therefore to develop teaching approaches which meet their needs. Laing (1999) spoke of the struggling learner claiming to know everything out of fear that they were ignorant even of what they were supposed to know, which adds another layer of complexity to the teacher's experience: the teacher can identify the struggling learner but requires nuanced skill in order to tease out specific areas of difficulty. The same is true methodologically for the researcher working with such participants. I have referred to my frontispiece quote that precision is not truth; in the search for Ryan's truth we must let go of any need for precision; there is

a strong methodological implication to explore how best to research and present cases such as Ryan's. The implications and importance of Ryan's inclusion in the current study is that researchers should not shy away from cases which are challenging: both this academic field and others will be poorer for their removal as outliers or contradictions.

Implications of the life capital question

Initially I had aimed for four case studies who were all monolingual English speakers: this was illustrative of the positivist mindset with which I began the research. I had wanted to reduce confounding variables despite my embracing the complexity approach. As I discussed in Methodology section 3.2.4, it turned out that two of my case studies were exposed to other languages at home despite no record of this on the school's systems.

Still, this inclusion of bilingualism has resulted in positive implications. I discussed life capital in section 10.3.4. of the discussion. It is important to acknowledge that bilingualism is the norm (Grosjean, 2010), and that very many young learners of L2s will come to their learning with knowledge of languages other than the language of schooling. Methodologically, then, this implies the importance of embracing these variables. Furthermore, the bicultural element of bilingualism has the potential to impact learning as I showed in Findings 7 and 9 and Discussion section 10.3.4.

It is important also to develop methods that further incorporate the concept of Life Capital (Consoli, 2022) into the CDST model, particularly from a qualitative point of view, and from a methodological perspective, further exploration of the Retrodictive Qualitative Modelling approach (Dörnyei, 2014) would merit further exploration.

11.5. Pedagogical implications

This section will address not only the implications of the present study for language teachers, but also implications for resource creation. In terms of concrete findings with regard to the participants' developmental trajectory of listening proficiency, the overriding story is one of influence of the L1 (largely English, but in the case of Fran I have also explored how Italian influenced her comprehension). Yet participants' trajectories have also been impacted by a wide range of individual differences, not all of which I had anticipated in my initial model comprising vocabulary, grammar and metacognitive skills, but were uncovered over the course of the study.

11.5.1. Listening to learn (reprise)

In section 11.2.2 I set out the theoretical argument indicating that for my participants, at least, listening to the language in order to further acquisition was at best unlikely and in many cases impossible, and showed in Figure 122 how my participants were rarely able to understand the 'syntax' section of Field's model of listening (2008b, 2013). In this way, 'input' could not become 'intake'. Field (2008:16) identified '*the idea that the listening lesson should serve to demonstrate recently taught language in everyday use.*' Sixteen years after publication of *Listening In The Language Classroom*, this idea remains the predominant rationale in text-book listening, by extension, justification for use of audio tracks by modern languages teachers in English secondary schools. Yet I have shown that in the first three years of secondary school, English learners of French lack the proficiency for auditory input to become intake. I stated in section 11.3 that teachers might have an overinflated concept of what beginner and intermediate learners are capable of during the listening process, and therefore a key pedagogical implication

must be an overhaul of the expectations teachers have of the purpose of listening in class and what can be achieved from it.

11.5.2. Learning to listen: Individual differences in the classroom

Field (2018) identified that the underlying challenge for the teaching of listening is,

‘The fact that listeners are individuals – they develop in different ways; they focus on different things... they have a way of developing their own techniques for handling uncertainties in the text.’

I have shown, for example, the many different aspects of life capital (Consoli, 2022) brought to the classroom: each student has different life capital, some of which the teacher can know and some of which they cannot. Another illustration could be Helen’s ‘self-subtitling’ approach, which presents a phenomenon inherent to the experience of (L2) listening which appears to have gone almost entirely unreported in the literature, yet is very unlikely to be unique to Helen.

Furthermore, one could take the finding of enormous flux between individual participants’ moving correlations of their linguistic systems. At a pedagogical level the finding demonstrates again the role of the individual’s complex dynamic system, where different elements might dominate at different times.

If, as I have argued, a discrepancy exists between the findings of case studies and group-level research, how should a teacher know what to teach? The answer might be as simple as to know one’s learners and to understand their individuality. Ducker (2022) suggests allocating students to classes that best meet their needs, although this suggestion may not be practical for many educational institutions, but also hints at a value in training students to be aware of the multiplicity of different factors that might be influencing their learning. This is a recommendation reinforced by Field (2018), who proposes that ‘*what*

we're going to have to do in future is to pass the initiative to a much greater degree to the listener'

There exists a debate about whether teaching language learning strategies is a better use of lesson time than teaching the language per se (Macaro, 2006), but if there is indeed a range of different needs among individuals in the classroom, addressing the importance of metacognition and students' taking responsibility for their own progress would have value in this regard. This will come about by in-depth, inclusive classroom discussion about how L2 listening works, the range of helpful approaches to tackle listening tasks, and a careful reflexive approach on the part of the teacher in order to empower listeners to use any tactics they have to hand which might be helpful.

11.5.3. Learning to listen: Teaching students with additional needs

The most recent data indicate that 18.4 per cent of students in England have a special educational need (Department for Education, 2024). Given Laing's (1999) point discussed in section 11.4.5 around students' difficulties in expressing both their knowledge and their ignorance, the implication that the teacher needs to know and understand the individual takes greater importance when it comes to those with additional needs. In the case of Ryan, a lower proficiency student who was diagnosed with special educational needs during the course of the study, I showed that he and I often might use a specific term differently: when Ryan told me at Time 3, *'I've figured it out, how to use my verbs,'* my instant response both as a teacher and as a doctoral researcher was to define this as use of conjugation in a range of tenses. Yet further probing resulted in the definition *'stuff how like you write I like and I hate'*. The Education Endowment Foundation (2018) lists approaches which it claims will help in closing the attainment gap, which are not specific to languages education. Howard (2023)

suggests a range of specific strategies to help language learners with special education needs while reiterating the need for an individual approach.

11.5.4. Learning to listen: Awareness of the influence of L1

I have discussed above that the present work is one of hypothesis-*forming*, not hypothesis-*testing*, and section 11.3 suggested that one key area of further research would be an exploration of the extent of the ‘phantom activation’ phenomenon, in which beginner listeners really do believe they have heard (sometimes unrelated) L1 words within the L2 speech stream. Until such research is carried out, I would offer a pedagogical recommendation for teachers to explore more deeply whether this is a phenomenon that their own students experience and build approaches to address or mitigate it.

11.5.5. Learning to listen: Implications for resource creation

I demonstrated in findings 1 that my participants consistently struggled with decoding skills. I found a number of ‘red herrings’ within the tracks which I had not anticipated would be problematic. Although L2 listening resources within the context of the English school system often do contain traps deliberately laid to catch the learner out (C. Bell et al., 2016), it is my impression that these red herrings were not such traps. To recap, some examples included:

- *J’adore le vélo et la musique*⁷⁸, (‘I love cycling and music’) whereby the word *musique* primed the listener to translate *vélo* as ‘violin’.

⁷⁸ Track 6 of test AS7 at T1

- *mais je voudrais bien une bouteille d'eau*⁷⁹ /mɛ ʒə vudʁɛ bjɛ̃ yn butɛj do/ ('but I'd quite like a bottle of water'), in which the context of ordering in a café primed listeners to translate *bouteille d'eau* /butɛj do/ as 'potato'.
- *Je ne peux pas rentrer seule*⁸⁰ /mɛ ʒə nə pø pa ʁɑ̃tʁɛ sœl/ ('but I am not allowed to come home alone'), in the context of life as a teenager, primed listeners to detect the word *parents* /paʁɑ̃/ within the *pas rent(rer)* /pa ʁɑ̃tʁɛ/.

Such issues point to the importance of piloting teaching resources before publication. The tracks had been designed for Comprehension Approach type testing, but my use of a paused translation approach uncovered different misunderstandings.

To illustrate both the inclination towards 'traps' in the current style of resources, and the power of the paused translation methodology to interrogate the efficacy of the Comprehension Approach, then, I revert to Time 17, at which point⁸¹ participants heard:

J'ai trouvé la glace à la vanille super. / À mon avis, le poulet rôti était froid, / alors je n'irai jamais dans ce café à l'avenir. / Les frites m'ont beaucoup plu / car elles n'étaient pas trop petites. / Pour moi, le gâteau au chocolat était bien / mais le prix était trop cher. / Selon moi, le sandwich au bœuf était vraiment savoureux.

(I found the vanilla ice-cream super / In my opinion, the roast chicken was cold / so I will never go to that café in the future / I really liked the chips / because they were not too small / For me the chocolate cake was good / but the price was too expensive / According to me, the beef sandwich was really tasty.)

⁷⁹ track 12, test C41, T16

⁸⁰ track 7 of test St3AP, T19

⁸¹ Test C11, at T17 only

The Comprehension Approach question to accompany this track can be seen in Figure 123; I have added the correct answers in red. The student would receive a mark out of four.

22 At the café

Listen to these people in a café. What are their opinions? Write **P** (positive), **N** (negative) or **P/N** (both positive and negative) in each box.

Example: Ice cream

1	Chicken	<input type="text" value="N"/>	(1 mark)
2	Chips	<input type="text" value="P"/>	(1 mark)
3	Chocolate cake	<input type="text" value="P/N"/>	(1 mark)
4	Beef sandwich	<input type="text" value="P"/>	(1 mark)

Figure 123: The Comprehension Approach set of questions accompanying the track C11, with correct answers in red

Analysis of the paused translation data suggests that only six participants would have got question one correct. Question two would have been overwhelmingly wrong, while most participants could identify *petites* ('small'), none could detect *pas trop* ('not too'). As for question three, most participants correctly detected both *chocolat* and *bien* to correctly identify the positive element of the statement, and 19 out of 64 also translated *mais* ('but'), which, with some level of inference, might have led them then to the suggestion that a negative counterpart was coming. Finally question four, which expected students to be able to understand the low-frequency false friend *savoureux* ('tasty') would have been answered correctly by only five participants.

I have shown that midstream function words are really not well identified, and nor are tenses, and question one hangs on two midstream function words: *trop* and *jamais*, as well as the statement *je n'irai jamais* ('I will never go'). The issue of the chips being *pas trop petites* also depends on the identification of a midstream function word. Might we conclude, therefore, that the reason that students declare listening difficult is not simply

because of the real time processing demands of making sense of a rapidly changing input out of their control, but also because the comprehension questions being asked of them contain a mixture of traps (e.g. *savoureux*) and developmentally highly demanding aspects such as rapidly identifying midstream function words which by their nature change the meaning of a sentence. The only discernible strategies that could be of use in this particular question is the use of ‘*mais*’ in the chocolate cake question, and this is largely a test-wise strategy rather than a listening strategy.

In fact, in the translation of *mais le prix était trop cher* (‘but the price was too expensive’), four participants alluded to the contextually relevant ‘sharing’. One higher proficiency participant translated *prix* as ‘taken’ (*pris* in French), which also indicates a lack of application of grammatical knowledge within this rapid-input environment, given that all the necessary precursors to ‘*pris*’ (e.g. a pronoun auxiliary) are absent, and the precursors to make this a noun (e.g. a definite article) are present.

I accept that by asking participants to translate all they could of this track, rather than to turn their attention to the comprehension questions, one might reasonably argue that they did not attend to the specificity of the questions given in Figure 123. Nonetheless, I have demonstrated, as Field (2008) stated, that meaning tends to be driven by content words, and the implications for resource creation should be that learning should therefore be developed and indeed tested primarily through content words, in the case of beginner and early intermediate learners. The present study adds to the body of evidence that illustrates the difficulties of function words, and the futility of assessing listeners through a Comprehension Approach resulting in data that are entirely meaningless in terms of telling either a learner, a teacher or a tester what the learner can or cannot do. In the place of Comprehension Approach exercises, resource writers (be they individual teachers or authors creating textbooks) should create listening activities to help develop decoding

skills (Graham & Santos, 2013; Leonard, 2019; Lange & Matthews, 2023) as well as listening skills and strategies more generally.

Teachers and resource-writers could create more appropriate resources using existing recordings in different fashions. For example:

- As demonstrated in the current thesis, giving paused translation exercises.
- Paused transcription exercises – ie dictation – but with the focus not on grammar and spelling, and instead on speech stream segmentation. The classic ‘Dictogloss’ activity is an excellent and low-preparation example of this.
- Modified paused transcription such gapped paused transcription, or faulty paused transcription, whereby the listener needs to identify (and correct) the error in the transcription.
- Elicited repetition exercises whereby the teacher plays a passage, stops and asks a student to repeat the last few words that they have heard, verbatim.
- Faulty echo exercises, whereby a teacher says a sentence twice, with one or two words differing between the first and second delivery. Students need to identify the difference between the two sentences.

11.6. Policy implications

The sections above addressing the pedagogical implications of learning to listen, then, have wider implications from a policy perspective, particularly with regard to assessment design. A new GCSE languages exam for first testing in May 2026 has introduced dictation into the listening exam (Department for Education, 2022a), and with this comes a certain degree of washback into classroom practice. While this is a welcome

development from my own perspective both academically and professionally, the present study suggests that it might not go far enough, for two reasons.

Firstly, only 20 per cent of marks in the listening paper are awarded for dictation (AQA, 2023); the remaining 80 per cent continue to be tested via a typical Comprehension Approach – an approach whose validity I have consistently questioned.

Secondly, I have shown throughout the present study that a paused translation approach might be a more valid way in which to assess listening proficiency than a paused transcription (i.e. dictation). I have also demonstrated that this can be marked with validity and reliability, and I would urge policymakers and the creators of examinations to explore its potential on a wider scale.

11.7. Conclusion

This chapter has addressed a range of implications of the present study. It has demonstrated its potential for developing further research ideas as well as indicated a range of key pedagogical implications to address some of the issues connected with listening in the language classroom. I will now move on to discussing the limitations of the research in the following chapter.

12. Limitations

No research project is without its limitations. Despite the exciting findings and implications I have covered in the previous chapters, the present study is also limited in a variety of ways.

12.1. Methodological limitations

My research design was not without its challenges and its limitations. Here I address some of the concerns I had arising from my methodology.

12.1.1. Panel conditioning

Panel conditioning, given the regular iterations of the test battery, presents a limitation. There are two chief concerns here: firstly, for those who were also case studies, it was possible that the very act of being interviewed about their listening skills, including stimulated recall protocols, seven times a year, might encourage this small sub-set of participants to reflect more deeply on their learning, as well as giving them a small amount of additional linguistic and metacognitive input. Indeed, case study Helen alluded to the learning she gleaned from the surveys on several occasions, although the other case study participants reported that they forgot their special role in the research between interviews. In order to best avoid confounding data caused directly by the interview process, stimulated recalls only ever followed listening passages which were not anchor points – in other words, to which the case study participants would not be exposed a second time within a data collection context. All this said, the CDST viewpoint accepts that all language use contributes to a learner's development, and with every additional use (data collection points included), a participant's language system is altered (De Bot

& Larsen-Freeman, 2011), and therefore one might argue that this is less a limitation, and more simply a by-product of linguistic research.

Secondly, there is a risk of whole-cohort panel conditioning with regard to the metacognitive listening questionnaire (MALQ) – it is possible that a canny participant might infer hints about best practice into the questions posed. Again, Helen recognised this and volunteered to me during interview that she had made use of the ideas suggested in the MALQ in order to further her listening skills. Other case-study participants appeared to give it no heed. I did not ask the full cohort whether they felt that the MALQ had given them ideas in how to improve their listening strategies, for fear of planting this as a very strategy in their head.

12.1.2. Access to equipment in school

The school's timetabling constraints represented a limitation. In an ideal world, all tests would have been equally spaced, held in an optimum timetable slot to ensure maximum concentration and performance. Yet demands such as access to computer rooms and set of 30 headphones, as well as the timing of the students' French lessons, meant that a more pragmatic approach had to be taken. This represents a theoretical limitation when compared with the ideal but is the reality of real-life data collection.

12.1.3. Linguistic tests

A perfect test design would have included 60 internationally recognised tests which had already been statistically checked and found to be of equal difficulty, without floor or ceiling effects. This would have allowed me to avoid using a Rasch model.

Unfortunately, even the DELF exams (section 3.3.2) only offer six tests which claim to be of equal difficulty. At the very least, my choice of tests shapes the data, and could be said to represent a limitation.

As the data collection continued, I increasingly wondered whether some participants were even *actually listening* to the tracks. In some cases, I could ascertain this (see section 3.8.2 on test fatigue and attrition). In other cases, I concluded that this is a limitation inherent in the testing of receptive skills: where a lack of understanding is demonstrated by a participant, it is impossible to know how much the participant has attempted to understand.

Further limitations related only to listening include the lack of control over some background noise in the test setting (for example, silly classroom behaviour while data are being collected, students in the corridor, interruptions by staff), or, on occasion, technical difficulties arising through headphone or computer use (e.g. participant knocks the headphone wire, or the computer gives an alert sound). The ephemeral nature of listening is a particular victim to these kinds of unpredictable events. These could have been mitigated in two ways: collecting data in a one-to-one setting whereby I could make a judgement on when an additional listen would be deserved, or setting up the initial test battery so participants could control how many times they listened to the track. Both of these mitigations have disadvantages, though: one-to-one for each participant simply would not have been practical for 80 participants, and participants' control over how many times they listened would have introduced a clear confounding variable into the data.

12.1.4. Use of a single school

Some might argue that situating the research within a single school represents a limitation, and even with data from 80 participants, caution must be taken not to extend findings to come to conclusions about to how all English secondary school pupils learn French. This said, CDST, with its acknowledgement of the profound role of

interconnectedness, resists the implication that any findings can be applied beyond the site and data in which the research has taken place (De Bot & Larsen-Freeman, 2011); instead it generalises to theory rather than to population.

There is also a benefit of situating all participants in a single research site whereby additional possible confounds are removed, such as the reputation of French within the institution, teaching styles, timetabling, or class groupings.

12.1.5. Use of my own school

Due to the volume of the data, I began marking the listening proficiency data before the end of the data collection period. I got to know many of the participants, not only through the marking, but also through my responsibilities as a staff member in the school.

Although I was not consciously influenced by this when marking or analysing data, it is quite conceivable that unconscious bias crept in.

However, as an anonymous researcher going into a school, my research would have been influenced too – for example the difficulty of the administrative task would have probably led to more missing data. Participants' reactions to an unknown researcher would not have been better, only different. An alternative scenario might have been to be a full-time researcher embedded in the school, whereby I could access the administrative side of the research easily (e.g. book rooms, find students for interviews etc) but separate from the languages department and visibly not a teacher, although without the behaviour management skills essential for ensuring quiet in the classroom during data collection sessions, other problems would have arisen.

12.2. Analytical limitations

12.2.1. Conceiving of the model

The width of the second language acquisition field demonstrates the enormous number of variables and their interactions that shape language development. The upshot of this is that I could have formed the boundaries of my model in a huge variety of ways – and indeed even the other research into listening development from a CDST perspective demonstrates the possibilities which I could have explored. Nonetheless, I could have designed my model in different fashions. Even with the existing data there remains the possibility of forming and exploring different models given the (over)ambition of my original research design.

For example, in the current thesis it is unclear as to how different teachers involved may have affected the patterns that have emerged from the findings. There were two constraints to this: one practical and one ethical.

From the practical perspective, there were four teaching groups for each of the three academic years, and the student groupings changed from one year to the next, which meant that any student could have fallen into one of 12 possible configurations of teaching group. Preliminary exploration of this presented further confounding variables, for example prior attainment levels were not even among the 12 groups, and group sizes were not equal (recall that I removed a number of participants at the outset due to reduced participation), but with an average of six or seven participants per group. Given these constraints, exploration of the influence of group membership seemed to me to lack academic rigour.

For ethical reasons I was also reluctant to explore group membership too deeply, as I felt I had a responsibility to my colleagues who were teaching the classes which I was

researching, and this was further complicated by the covid lockdowns, which had strong practical and emotional effects on some of them.

When analysing the data and researching the discussion chapter, it also became clear to me that there was potential to explore the relationships of some of the metacognitive subsystems with those of some of the linguistic subsystems: for example I began to wonder what the relationships might have been between Mental Translation and lexical knowledge or Directed Attention and grammatical knowledge. Again, the data exist so it would be possible to explore this in subsequent projects.

12.2.2. Correlation windows

Bulté & Housen (2020) state that the size of the moving correlation windows might have an impact on our findings. Therefore my decision to create windows of three time points for my moving correlation represents a limitation: a different direction of research (and an interesting methodological paper) might be to explore the differences in findings with different window sizes: for the data in present study I would suggest comparing a moving window of three time points with that of six or seven time points, which would equate to comparing learning on a termly basis with learning on an annual basis.

13. Conclusion

On my opening page I showed a tiled piece of graffiti art I had found in rue Lhomond in Paris in 2022, which stated *exactitude n'est pas vérité*. The incorrect French grammar was pleasingly ironic in a street dedicated to a French educator and grammarian, and illustrated the inherent tension between precision and truth.

The present study set out to explore the developmental trajectory of second language listening among a group of English secondary school students learning French. As an experienced teacher, the Complex Dynamic Systems approach made intuitive sense to me from the moment I discovered it. I began by asking whether different subsystems which contribute to second language listening would carry different levels of importance at different times in the learning journey, but I do not believe I have successfully answered that question. Instead I have shown that a group level, vocabulary carries more weight than grammar, but there were times when the two subsystems appeared to co-operate and other times when one appeared to grow at the expense of the other, which is a new contribution to knowledge. I also showed that the metacognitive systems roughly fell into two pairs, whereby Directed Attention and Person Knowledge worked in consort ('I am less anxious, and I pay better attention'), as did Problem-Solving and Mental Translation.

However, when exploring the same questions at an individual level, the experiences of Fran, Helen, Theo and Ryan bore very little resemblance to the group average. Instead, each of these individuals' experiences showed dramatic fluctuations and changes from time point to time point, and when looking at the paused translation data, even from moment to moment. Helen's fascinating self-subtitling approach is a key contribution to knowledge, as is the fluctuating role of life capital (for example how Italian is a help and hindrance to Fran's learning at different points). Ryan's stumble through his learning of

French, struggling with additional needs and remote learning, is also an important contribution to our pedagogical knowledge of this kind of under-researched student.

I showed that classroom research is messy – particularly classroom research into students aged 11 to 14, and particularly when a pandemic is thrown into the mix too. I showed the robustness of the CDST approach to deal with all manner of unpredictable elements, and I will continue to advocate for a move against the positivist urge to keep research tidy or precise, and to remove difficult outliers or unusual cases. Had I followed a more positivist approach, the only case study who would have fitted the mould would have been Helen – Theo and Fran would have been ruled out for their ‘inconvenient’ bilingualism, and Ryan for his erratic answers and chaotic data, which would have been deemed ‘noise’ rather than sound.

Complex, dynamic systems are all around us at all times. I set out to report on one (which was in Chapter 5 / Findings 2), and in doing so I uncovered the moment-to-moment systems apparent in Chapter 4, and the individual systems of each of the case studies. My own journey through this doctorate was also a complex dynamic system, with moments of plateau (fighting through the Rasch modelling was particularly slow process and I was stuck in an ‘attractor’ state for some time) and other moments with real jumps in progress. Still, the use of the Rasch model also becomes an important contribution to methodological knowledge as a tool for measuring receptive skills over time by putting different scores onto a single scale.

Second language listening continues to operate as a complex dynamic system on many levels, with researchers conceiving of all manner of subsystems which can contribute to it. Field’s model has dominated this area of research for nearly 20 years, but the present study has made the important contribution to knowledge that not each of his levels of

analysis carry the same weight throughout the listening process, or even for each individual listener. This proposition becomes even more powerful by overlaying the concept of the fuzzy lexical representation, driven as it is by the omnipresence of the first language in the second language listening experience, and justifies the contribution of the paused translation methodology as opposed to the dominant paused transcription approach

Finally, and to my mind most importantly, I have also demonstrated the importance of the individual and their range of experiences when they are listening to a second language. After all, if listening takes place in the unique and wonderful (and sometimes slightly terrifying) place that is the human mind, each listening experience will be personal – and precision cannot be truth.

14. References

- Akkerman, S., & Bakker, A. (2011). Boundary Crossing and Boundary Objects. *Review of Educational Research*, 81, 132–169.
- Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, 36(1), 94–106. <https://doi.org/10.1016/j.system.2007.12.002>
- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35, 79–113.
- Allan, D. (1995). Oxford Placement Test (French). Retrieved from <https://www.lang.ox.ac.uk/french#collapse663211>
- Anderson, J. R. (1990). *Cognitive psychology and its implications*. New York: WH Freeman.
- Andrews, J., Robinson, D., & Hutchinson, J. (2017). *Closing the Gap? Trends in Educational Attainment and Disadvantage*. Retrieved from <https://epi.org.uk/publications-and-research/closing-gap-trends-educational-attainment-disadvantage/>
- Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. <https://doi.org/10.1007/978-981-13-7496-8>
- Andringa, S., & Godfroid, A. (2020). Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. <https://doi.org/10.1017/S0267190520000033>
- Andringa, S., Olsthoorn, N., Van Beuningen, C. G., Schoonen, R., & Hulstijn, J. (2012). Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach. *Language Learning*, 62(SUPPL. 2), 49–78. <https://doi.org/10.1111/j.1467-9922.2012.00706.x>
- Anzaki, F. (2021). Differences between Conduction Aphasia and Wernicke’s Aphasia. *The Tokai Journal of Experimental and Clinical Medicine*, 26(2), 45–61.
- AQA. (2016). *GCSE French (8658)*. Manchester.
- AQA. (2023). GCSE for French: Specification for teaching from September 2024 onwards; For GCSE exams in 2026 onwards. Retrieved November 23, 2024, from <https://cdn.sanity.io/files/p28bar15/green/6c81ef5090708cb0792247904e02e6b0c40f394a.pdf?se=french-8652/specification>
- Assessment and Qualifications Alliance. (2018). GCSE French Specification Specification for first teaching in 2016. Retrieved January 5, 2018, from <https://www.aqa.org.uk/subjects/languages/gcse/french-8658>
- Baba, K. (2020). Exploring dynamic developmental trajectories of writing fluency. In G. G. Fogal & M. Verspoor (Eds.), *Complex Dynamic Systems Theory and L2 Writing Development* (pp. 3–26). Amsterdam: John Benjamins.

- Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning*, 64(1), 1–35. <https://doi.org/10.1111/lang.12033>
- Balestra, S., Eugster, B., & Liebert, H. (2020). Summer-born struggle: The effect of school starting age on health, education, and work. *Health Economics (United Kingdom)*, 29(5), 591–607. <https://doi.org/10.1002/hec.4005>
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148.
- Banegas, D. L., & Consoli, S. (2024). Practitioner research in education : A call for a reality check Practitioner research in education : A call for a reality check ABSTRACT. *International Journal of Research & Method in Education*, 1–14. <https://doi.org/10.1080/1743727X.2024.2378708>
- Baxter, P., Bekkering, H., Dijkstra, T., Droop, M., van den Hurk, M., & Leoné, F. (2022). Constrasting orthographically similar words facilitates adult second language vocabulary learning. *Learning and Instruction*, 80, 1–20.
- Becker, S. (2015). Metacognitive instruction in L2 French: An analysis of listening performance and automaticity. *ProQuest Dissertations and Theses*, 186. Retrieved from https://search.proquest.com/docview/1712677721?accountid=10673%0Ahttp://openurl.ac.uk/redirect/athens:edu/?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&a
- Beglar, D., & Nation, P. (2014). Assessing vocabulary. In A. J. Kunnan (Ed.), *The Companion to Language Assessment* (pp. 173–184). Chichester: Wiley Blackwell.
- Bell, C., McLachlan, A., & Ramage, G. (2016). *Studio AQA GCSE French*. London: Pearson Education Ltd.
- Bell, J. F., & Daniels, S. (1990). Are Summer-born Children Disadvantaged? The Birthdate Effect in Education. *Oxford Review of Education*, 16(1), 67–80. <https://doi.org/10.1080/0305498900160106>
- Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. J. (2010). What makes listening difficult? *University Of Maryland*, (April), 92. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a550176.pdf>
- Board, K., & Tinsley, T. (2016). *Language Trends survey*. Reading: CfBT Education Trust.
- Bond, Z. S. (1999). *Slips of the ear: Errors in the perception of casual conversation*. Cambridge, MA: Academic Press.
- Bonk, W. J. (2000). Second Language Lexical Knowledge and Listening Comprehension. *International Journal of Listening*, 14(1), 14–31. <https://doi.org/10.1080/10904018.2000.10499033>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch Analyses in the Human Sciences*. *Advances in Rasch Analyses in the Human Sciences*. <https://doi.org/10.1007/978-3-030-43420-5>

- Boone, W. J., Staver, J. R., & Yale, M. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer US.
- Bozorgian, H. (2012). Metacognitive Instruction Does Improve Listening Comprehension. *ISRN Education, 2012*, 1–6. <https://doi.org/10.5402/2012/734085>
- Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 1*–25. <https://doi.org/https://doi.org/10.1080/14780887.2020.1769238>
- Brenders, P., van Hell, J., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109*(4), 383–396.
- British Council. (2023). *Language Trends 2023*. Retrieved from https://internationalnewcastle.org.uk/wp-content/uploads/2023/06/Language-Trends-England_2023-Report-June-2023.pdf
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 26–34.
- Brunfaut, T., & Révész, A. (2015). The Role of Task and Listener Characteristics in Second Language Listening. *TESOL Quarterly, 49*(1). <https://doi.org/10.1002/tesq.168>
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Bulté, B., & Housen, A. (2020). A DUB-Inspired Case Study of Multidimensional L2 Complexity Development: Competing or Connecting Growers? In W. Lowie, M. Michel, M. Keijzer, & R. Steinkrauss (Eds.), *Usage-Based Dynamics in Second Language Development* (pp. 50–86). Hoboken: Wiley. <https://doi.org/10.21832/9781788925259-006>
- Byrne, D. (2021). A worked example of Braun and Clarke’s approach to reflexive thematic analysis. *Qual Quant, 56*, 1391–1412.
- Caelli, K., Ray, L., & Mill, J. (2003). 2003. Clear as Mud: Toward Greater Clarity in Generic Qualitative Research. *International Journal of Qualitative Methods, 2*.
- Cao, X., Yamashita, N., & Ishida, T. (2016). How non-native speakers perceive listening comprehension problems: Implications for adaptive support technologies. *Communications in Computer and Information Science, 647*, 89–104. https://doi.org/10.1007/978-981-10-2618-8_8
- Carney, N. (2021). Diagnosing L2 listeners’ difficulty comprehending known lexis. *TESOL Quarterly, 55*(2), 536–567.
- Carroll, S. E. (2004). Segmentation: Learning how to “hear words” in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227–254. <https://doi.org/10.1111/j.0079-1636.2004.00136.x>
- Caspi, T., & Lowie, W. (2013). The dynamics of L2 vocabulary development : a case study of receptive and productive knowledge A dinâmica do desenvolvimento do vocabulário L2 : um. *Revista Brasileira de Linguística Aplicada, 1*–15. <https://doi.org/10.1590/S1984-63982013005000002>

- Cattan, S., Farquharson, C., Krutikova, S., Phimister, A., Salisbury, A., & Sevilla, A. (2021). *Home Learning Experiences Through The Covid-19 Pandemic*. Retrieved from <https://ifs.org.uk/publications/home-learning-experiences-through-covid-19-pandemic>
- Chang, A. C. S., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, 68(1), 31–40. <https://doi.org/10.1093/elt/cct052>
- Chang, P. (2017). *A complex dynamic systems theory perspective on learner individual differences in second language listening development*. Auckland.
- Chang, P., & Zhang, L. J. (2021). A CDST Perspective on Variability in Foreign Language Learners' Listening Development. *Frontiers in Psychology*, 12(February), 1–17. <https://doi.org/10.3389/fpsyg.2021.601962>
- Chang, P., & Zhang, L. J. (2023). Learners' Listening Development. In *Language Aptitude Theory and Practice* (pp. 302–325). Cambridge: Cambridge University Press. <https://doi.org/https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1017/9781009076463>
- Chen, A. (2009). Listening strategy instruction: Exploring Taiwanese college students' strategy development. *Asian EFL Journal*, 11(2), 54–85.
- Chen, A. (2013). EFL listeners' strategy development and listening problems: A process-based study. *Journal of Asia TEFL*, 10(3), 81–101. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84887918944&partnerID=40&md5=bc3d1907c4a5b6fad3164abbc4b4827>
- Chenjun, D., & Li, L. (2012). The Effectiveness of Explicit Instruction of Certain Decoding Skills in Improving Chinese EFL Listeners' General Comprehension Performance. *Chinese Journal of Applied Linguistics*, 35(2), 243–256. <https://doi.org/10.1515/cjal-2012-0017>
- Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65.
- Collen, I. (2021). *Language Trends*. Retrieved from https://www.britishcouncil.org/sites/default/files/language_trends_2021_report.pdf
- Consoli, S. (2021a). Critical incidents in a teacher-researcher and student-participant relationship: What risks can we take? In C. Cunningham & C. J. Hall (Eds.), *Vulnerabilities, challenges and risks in applied linguistics* (pp. 147–161). Bristol: Multilingual Matters.
- Consoli, S. (2021b). Understanding Motivation through Ecological Research: The Case of Exploratory Practice. In R. Sampson & R. Pinner (Eds.), *Complexity Perspectives on Researching Language Learner and Teacher Psychology* (pp. 120–135). Bristol: Multilingual Matters.
- Consoli, S. (2022). Life Capital: An Epistemic and Methodological Lens for TESOL Research. *TESOL Quarterly*, 0(0), 1–13. <https://doi.org/10.1002/tesq.3154>
- Consoli, S., & Ganassin, S. (2022). *Reflexivity in Applied Linguistics: Opportunities, Challenges, and Suggestions*. Abingdon: Routledge.

- Cook, S. V. (2012). *Phonological Form in L2 Lexical Access: Friend or Foe?* University of Maryland.
- Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology, 7*, 1–17.
- Cross, J. (2009). Diagnosing the process, text, and intrusion problems responsible for L2 listeners' decoding errors. *Asian EFL Journal, 11*(2), 31–53.
- Cross, J. (2010). Raising L2 listeners' metacognitive awareness: A sociocultural theory perspective. *Language Awareness, 19*(4), 281–297. <https://doi.org/10.1080/09658416.2010.519033>
- Cross, J., & Vandergrift, L. (2015). Guidelines for designing and conducting L2 listening studies. *ELT Journal, 69*(1), 86–89. <https://doi.org/10.1093/elt/ccu035>
- Cutler, A. (2002). Listening to a second language through the ears of a first. *Interpreting, 5*(1), 1–23. <https://doi.org/10.1075/intp.5.1.02cut>
- Cutler, A., & Clifton, C. (1999). Comprehending spoken language, a blueprint of the listener. In *The neurocognition of language*. Oxford: Oxford University Press.
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269–284.
- Czerniawski, G. (2022). Power, positionality and practitioner research: schoolteachers' experiences of professional doctorates in education. In *AERA 2022 Conference*.
- De Bot, K., & Larsen-Freeman, D. (2011). Researching SLD from a DST perspective. In M. Verspoor, K. De Bot, & W. Lowie (Eds.), *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 5–28). Amsterdam: John Benjamins Publishing Company.
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(01), 7. <https://doi.org/10.1017/S1366728906002732>
- Deckner, S. E. (2014). ISSUES OF PUPIL MOTIVATION FOR LANGUAGE LEARNING IN YEAR 7 by, (October).
- Department for Education. (2019). Schools, pupils and their characteristics: January 2019, (January), 1–15. Retrieved from www.nationalarchives.gov.uk/doc/open-government-licence/version/3
- Department for Education. (2020). Special educational needs in England. Retrieved July 14, 2024, from <https://explore-education-statistics.service.gov.uk/findstatistics/special-educational-needs-in-england#releaseHeadlines-ta>
- Department for Education. (2021). Understanding Progress in the 2020/21 Academic Year – Findings from the summer term and summary of all previous findings'.
- Department for Education. (2022a). Reforms to encourage more students to take up language GCSEs. Retrieved November 23, 2024, from <https://www.gov.uk/government/news/reforms-to-encourage-more-students-to-take->

up-language-gcses

- Department for Education. (2022b). Understanding progress in the 2020/21 Academic Year: Extension report covering the first half of the autumn term 2021/22.
- Department for Education. (2023). The National Curriculum. Retrieved October 30, 2023, from <https://www.gov.uk/national-curriculum/key-stage-3-and-4>
- Department for Education. (2024). Special Educational Needs in England. Retrieved July 19, 2024, from <https://explore-education-statistics.service.gov.uk/find-statistics/special-educational-needs-in-england>
- Dijkstra, T. (2009). The multilingual lexicon. In D. Sandra, J.-O. Östman, & J. Verschueren (Eds.), *Cognition and language use: Handbook of pragmatics highlights* (pp. 369–388). Amsterdam: Benjamins.
- Dong, J. (2016). A dynamic systems theory approach to development of listening strategy use and listening performance. *System*, *63*, 149–165. <https://doi.org/10.1016/j.system.2016.10.004>
- Donnelly, R., & Patrinos, H. A. (2022). Learning loss during Covid-19: An early systematic review. *Prospects*, *51*, 601–609.
- Dörnyei, Z. (2014). Researching complex dynamic systems: “Retrodictive qualitative modelling” in the language classroom. *Language Teaching*, *47*(1), 80–91. <https://doi.org/10.1017/S0261444811000516>
- Douglas Fir Group. (2016). A Transdisciplinary Framework for SLA in a Multilingual World. *Modern Language Journal*, *100*, 19–47. <https://doi.org/10.1111/modl.12301>
- Ducker, N. (2022). A situated examination of the complex dynamic nature of L2 listening in EMI. *Language Teaching Research*, *0*(0), 1–51.
- Ducker, N. (2024). Capturing the situated, dynamic nature of EMI-lecture listening comprehension in real time. *Journal of English Medium Instruction*, *32*, 185–212.
- Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*(1), 37–53.
- Dunkel, P., & Davis, J. (1994). The effects of rhetorical signalling cues on the recall of English lecture information by speakers of English as a native or second language. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 55–74). Cambridge: Cambridge University Press.
- Duyck, W. (2005). Translation and Associative Priming With Cross-Lingual Pseudohomophones: Evidence for Nonselective Phonological Activation in Bilinguals. *Journal of Experimental Psychology*, *31*(6), 1340–1359.
- Eastman, J. (1991). Learning to listen and to comprehend: the beginning stages. *System*, *19*(3), 179–187.
- Education Endowment Foundation. (2018). *The Attainment Gap*. Retrieved from https://d2tic4wv01iusb.cloudfront.net/production/documents/support-for-schools/bitesize-support/EEF_Attainment_Gap_Report_2018.pdf?v=1714301618
- Education Endowment Foundation. (2022). *The Impact of COVID-19 on Learning: A review of the evidence*. Retrieved from

https://d2tic4wv01iusb.cloudfront.net/production/documents/guidance-for-teachers/covid-19/Impact_of_Covid_on_Learning.pdf?v=1714234804

- Ellis, N. (2017). Salience in usage-based SLA. In S. Gass, P. Spinner, & J. Behney (Eds.), *In Salience in second language acquisition* (pp. 21–40). New York: Routledge.
- Ellis, N. C., & Larsen-Freeman, D. (2006). Language Emergence: Implications for Applied Linguistics--Introduction to the Special Issue. *Applied Linguistics*, 27(4), 558–589. <https://doi.org/10.1093/applin/aml028>
- Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). Second language verb-argument constructions are sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism*, 4(4), 405–431. article. <https://doi.org/10.1075/lab.4.4.01ell>
- Engzell, P., Frey, A., & Verhagen, M. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17). Retrieved from <https://doi.org/10.1073/pnas.2022376118>
- Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36, 407–422.
- Estes, R. L. (2014). *Lexical segmentation in L2 Spanish listening*. University of California.
- Fan, J., & Bond, T. (2019). Applying Rasch measurement in language assessment: Unidimensionality and local independence. In *Quantitative Data Analysis for Language Assessment Volume 1* (pp. 83–102).
- Field, J. (1998). Skills and strategies : towards a new methodology for listening. *ELT Journal*, 52(2), 110–118. <https://doi.org/10.1093/elt/52.2.110>
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57(5), 325–334. <https://doi.org/10.1093/elt/57.4.325>
- Field, J. (2008a). Bricks or Mortar: Which Parts of the Input Does a Second Language Listener Rely On? *TESOL Quarterly*, 42(3), 411–432. <https://doi.org/10.1002/j.1545-7249.2008.tb00139.x>
- Field, J. (2008b). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Field, J. (2008c). Revising segmentation hypotheses in first and second language listening. *System*, 36(1), 35–51. <https://doi.org/10.1016/j.system.2007.10.003>
- Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes*, 10(2), 102–112. <https://doi.org/10.1016/j.jeap.2011.04.002>
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening*. Cambridge: Cambridge University Press.
- Field, J. (2018). An Interview with John Field. Retrieved July 19, 2024, from <https://teacherdevelopmentlab.wordpress.com/2018/02/21/an-interview-with-john-field/>
- Field, J. (2019). *Rethinking the second language listening test: from theory to practice*.

Sheffield: Equinox.

- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24, 369–383.
- Fogal, G. G., & Verspoor, M. (2020). *Complex dynamic systems theory and L2 writing development*. Amsterdam ; Philadelphia.
- Galloway, N. (2016). Researching your own students. In *Doing Research in Applied Linguistics: Realities, Challenges and Solutions* (pp. 146–156). Abingdon: Routledge.
- Gass, S. (1994). The reliability of second-language grammaticality judgements. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 302–322). Hillsdale, NJ: Erlbaum.
- Godfroid, A., & Andringa, S. (2023). Uncovering Sampling Biases, Advancing Inclusivity, and Rethinking Theoretical Accounts in Second Language Acquisition: Introduction to the Special Issue SLA for All? *Language Learning*, 73(4), 981–1002. <https://doi.org/10.1111/lang.12620>
- Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75. [https://doi.org/10.1016/S0346-251X\(99\)00060-3](https://doi.org/10.1016/S0346-251X(99)00060-3)
- Goh, C., & Kaur, K. (2017). Insights into young learners' metacognitive awareness about listening. *The European Journal of Applied Linguistics and TEFL*, 2(1), 5–26.
- Goh, C., & Taib, Y. (2006). Metacognitive instruction in listening for young learners. *ELT Journal*, 60(3), 222–232. <https://doi.org/10.1093/elt/ccl002>
- Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy Lexical Representations in Adult Second Language Speakers. *Frontiers in Psychology*, 12(November). <https://doi.org/10.3389/fpsyg.2021.732030>
- Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, 34(2), 165–182. <https://doi.org/10.1016/j.system.2005.11.001>
- Graham, S. (2007). Learner strategies and self-efficacy: Making the connection. *Language Learning Journal*, 35(1), 81–93. <https://doi.org/10.1080/09571730701315832>
- Graham, S., & Macaro, E. (2007). Designing year 12 strategy training in listening and writing: From theory to practice. *Language Learning Journal*, 35(2), 153–173. <https://doi.org/10.1080/09571730701599203>
- Graham, S., & Macaro, E. (2008). Strategy instruction in listening for low-intermediate learners of French. *Language Learning*, 58(4), 7470783.
- Graham, S., & Santos, D. (2013). Selective listening in L2 learners of French. *Language Awareness*, 22(1), 56–75. <https://doi.org/10.1080/09658416.2011.652634>
- Graham, S., Santos, D., & Vanderplank, R. (2008). Listening comprehension and strategy use: A longitudinal exploration. *System*, 36(1), 52–68. <https://doi.org/10.1016/j.system.2007.11.001>
- Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning*

- and Teaching*, 4(1), 1–20. <https://doi.org/10.1080/17501220802385866>
- Graham, S., Santos, D., & Vanderplank, R. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching Research*, 15(4), 435–456. <https://doi.org/10.1177/1362168811412026>
- Graves, M. (2009). *Essential readings on vocabulary instruction*. Newark, DE: International Reading Association.
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, 38(4), 299–310. <https://doi.org/10.3758/BF03207159>
- Grosjean, F. (2010). *Bilingual: life and reality*. Cambridge, MA: Harvard University Press.
- Güvercin, D., Kesici, A. E., & Akbaşlı, S. (2022). Distance Education Experiences of Teacher-Parents during the COVID-19. *Athens Journal of Education*, 9(4), 593–614. <https://doi.org/10.30958/aje.9-4-4>
- Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL - International Review of Applied Linguistics in Language Teaching*, 40, 69–87.
- Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing*, 40(March), 44–59. <https://doi.org/10.1016/j.jslw.2018.03.001>
- Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, 17(3), 798–810. <https://doi.org/10.1016/j.concog.2007.12.006>
- Hepford, E. (2020). The elusive phase shift Capturing changes in L2 writing development and interaction between the cognitive and social ecosystems. In G. G. Fogal & M. Verspoor (Eds.), *Complex Dynamic Systems Theory and L2 Writing Development* (pp. 161–182). Amsterdam: John Benjamins Publishing Company.
- Hiver, P., & Al-hoorie, A. H. (2022). Transdisciplinary research methods and complexity theory in applied linguistics : introduction to the special issue, 60(1), 1–6.
- Hiver, P., & Al-Hoorie, A. H. (2020). *Research Methods for Complexity Theory in Applied Linguistics*. Bristol: Multilingual Matters.
- Howard, K. B. (2023). Supporting learners with special educational needs and disabilities in the foreign languages classroom. *Support Learn*, 38, 154–161.
- Hui, B., & Godfroid, A. (2020). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, 1–27. <https://doi.org/10.1017/S0142716420000193>
- Hulstijn, J. H. (2015). Discussion: How Different Can Perspectives on L2 Development Be? *Language Learning*, 65(1), 210–232. <https://doi.org/10.1111/lang.12096>
- Jordan, G. (2004). *Theory Construction in Second Language Acquisition*. Amsterdam: John Benjamins.
- Joy, R., & Murphy, E. (2012). Inclusion of Children with Special Educational Needs in

- an Intensive French as a Second Language Program: From Theory to Practice. *Canadian Journal of Education*, 35(1), 102–119.
- Kennedy, S., & Blanchet, J. (2014). Language awareness and perception of connected speech in a second language. *Language Awareness*, 23(1–2), 92–106. <https://doi.org/10.1080/09658416.2013.863904>
- Kennedy, S., Guénette, D., Murphy, J., & Allard, S. (2015). Le rôle de la prononciation dans l'intercompréhension entre locuteurs de français lingua franca. *Canadian Modern Language Review*, 71(1), 1–25. <https://doi.org/10.3138/cmlr.2139>
- Khaing, T. O. (2023). An EFL Learner's Development of Metacognitive Listening Strategy Use and Listening Performance: A Complex Dynamic Systems Theory Approach. In A. M. Wind & B. Dóczy (Eds.), *DEAL 2023: Multiple Perspectives in English Applied Linguistics* (pp. 119–148). Budapest: Eötvös University Press. <https://doi.org/10.21862/ELTE.DEAL.2023>
- Khoshsima, H., Salehzadeh, M., & Saed, A. (2011). On the Cognitive Problems of Nautical Students' Listening Comprehension. *The International Maritime English Conference*, (March 2012), 200–211. Retrieved from <http://imla-imec.com/IMEC23Proceedings.pdf>
- Klimova, B. (2021). An Insight into Online Foreign Language Learning and Teaching in the Era of COVID-19 Pandemic. *Procedia Computer Science*, 192, 1787–1794.
- Knight, B. (2022). *Complex Adaptive Behaviours in Small Group Interaction: A Year 4 Classroom Case Study Of Learning As "Emergence."* University of the West of England.
- Kostoulas, A., Stelma, J., Mercer, S., Cameron, L., & Dawson, S. (2017). Complex Systems Theory as a Shared Discourse Space for TESOL. *TESOL Journal*, 1–15. <https://doi.org/10.1002/tesj.317>
- Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. London: Longman.
- Laing, R. D. (1999). *Knots*. London: Routledge.
- Lange, K. (2018). Analyzing difficulties in aural word recognition for Japanese English learners: Identifying function words in connected speech. *CASELE Research Bulletin*, 48, 63–73.
- Lange, K., & Matthews, J. (2023). Analyzing trends in the aural decoding errors of Japanese EFL learners. *IRAL - International Review of Applied Linguistics in Language Teaching*, 61(2), 545–570.
- Lanvers, U. (2020). Changing language mindsets about modern languages: a school intervention. *Language Learning Journal*, 48(5), 571–597. <https://doi.org/10.1080/09571736.2020.1802771>
- Lanvers, U., & Coleman, J. A. (2017). The UK language learning crisis in the public media: a critical analysis. *The Language Learning Journal*, 45(1), 3–25. <https://doi.org/10.1080/09571736.2013.830639>
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*,

- 27(4), 590–619. <https://doi.org/10.1093/applin/aml029>
- Larsen-Freeman, D. (2017). Complexity Theory. In *Complexity Theory And Language Development* (pp. 11–50). Amsterdam: John Benjamins Publishing Company.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.
- Leclercq, P., & Edmonds, A. (2014). How to assess L2 proficiency? An overview of proficiency assessment research. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp. 3–23). Bristol: Multilingual Matters.
- Leonard, K. R. (2019). Examining the relationship between decoding and comprehension in L2 listening. *System*, 87, 102150. <https://doi.org/10.1016/j.system.2019.102150>
- Leonard, L. (1982). Phonological deficits in children with developmental language impairment. *Brain and Language*, 16(1), 73–86.
- Levak, N. (2013). *Using Second Life and Skype for Facilitating the Development of Listening Comprehension in Second Language Learners*. University of Southern Queensland.
- Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin and Review*, 18(6), 1035–1041. <https://doi.org/10.3758/s13423-011-0163-0>
- Linacre, J. M. (2012). Winsteps (version 3.74). available from <http://www.winsteps.com>.
- Liu, B., Wang, Z., & Jin, Z. (2009). The integration processing of the visual and auditory information in videos of real-world events: An ERP study. *Neuroscience Letters*, 461(1), 7–11. <https://doi.org/10.1016/j.neulet.2009.05.082>
- Liu, N. F. (2002). *Processing problems in l2 listening comprehension of university students in Hong Kong*. Hong Kong Polytechnic University.
- Long, D. R. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition*, 12(1), 65–80.
- Lotfi, G., Maftoon, P., & Birjandi, P. (2016). Learning to listen: does intervention make a difference? *Language Learning Journal*, 44(1), 107–123. <https://doi.org/10.1080/09571736.2012.734320>
- Lowie, W. (2012). *Essential Statistics for Applied Linguistics*. London: Palgrave MacMillan.
- Lowie, W. (2013). Dynamic systems theory approaches to second language acquisition. In C. Chappelle (Ed.), *The Encyclopaedia of Applied Linguistics*. Chichester: Wiley.
- Lowie, W. (2023). Significance in Moving Correlations.
- Lowie, W., & Verspoor, M. (2015). Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. *Language Learning*, 65(1), 63–88. <https://doi.org/10.1111/lang.12093>
- Lowie, W., & Verspoor, M. (2018). Individual Differences and the Ergodicity Problem.

Language Learning.

- Luo, G., Seow, A., & Chin, C. L. (2001). Linking and Anchoring Techniques in Test Equating Using the Rasch Model. *Proceedings of the 5th CAA Conference*, 1–12. Retrieved from <https://dspace.lboro.ac.uk/2134/1817>
- Macaro, E. (2006). Strategies for language learning and for language use: revising the theoretical framework. *The Modern Language Journal*, 90(3), 320–337. <https://doi.org/10.1111/j.1540-4781.2006.00425.x>
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1). <https://doi.org/10.1017/S0261444817000350>
- MacIntyre, P. D., & Ducker, N. (2022). The idiodynamic method: A practical guide for researchers. *Research Methods in Applied Linguistics: A Practical Resource*, 1(2), 100007.
- MacIntyre, P. D., MacKay, E., Ross, J., & Abel, E. (2017). The emerging need for methods appropriate to study dynamic systems. In L. Ortega & Z. Han (Eds.), *Complexity Theory And Language Development* (pp. 97–122). Amsterdam: John Benjamins Publishing Company.
- Mackey, A., & Gass, S. (2005). *Second Language Research Methodology and Design*. New York: Routledge.
- Mady, C., & Muhling, S. (2017). Instructional Supports for Students with Special Education Needs in French as a Second Language Education: A Review of Canadian Empirical Literature. *Journal of Education and Learning*, 6(3), 14–22.
- Makhovych, I. (2024, February 27). What I've learnt from teaching in Kyiv amid a war. *Nature*. Retrieved from https://www.nature.com/articles/d41586-024-00588-6?utm_medium=affiliate&utm_source=commission_junction&utm_campaign=CONR_PF018_ECOM_GL_PBOOK_ALWYS_DEEPLINK&utm_content=textlink&utm_term=PID1612532&CJEVENT=2a2b2225057711ef82354d370a18b8f8&countryCode=de
- Mareschal, C. (2007). *Student perceptions of a self-regulatory approach to second language listening comprehension development*. University of Ottawa.
- Marin, M. M., & Peltzer-Karpf, A. (2009). Towards a Dynamic Systems Approach to the Development of Language and Music: Theoretical Foundations and Methodological Issues. *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland*, (Escom), 284–292.
- Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, 72, 23–36. <https://doi.org/10.1016/j.system.2017.10.005>
- Mavridi, S. (2022). *Language Teaching Experiences During Covid-19*. Retrieved from https://www.britishcouncil.org/sites/default/files/language_trends_2021_report.pdf
- McKinley, J. (2017). Overcoming problematic positionality and researcher objectivity. In *Doing Research in Applied Linguistics: Realities, Challenges and Solutions* (pp. 37–47). Abingdon: Routledge.

- McKinley, J., & Rose, H. (2019). *The Routledge Handbook of Research Methods in Applied Linguistics*. The Routledge Handbook of Research Methods in Applied Linguistics. Routledge. <https://doi.org/10.4324/9780367824471>
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.
- Mecarty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning, 11*, 323–348.
- Mercer, S. (2016). Complexity and language teaching. In *The Routledge Handbook of English Language Teaching* (pp. 473–485). Abingdon: Routledge.
- Meyer, J. P., & Zhu, S. (2013). Fair and Equitable Measurement of Student Learning in MOOCs: An Introduction to Item Response Theory, Scale Linking and Score Equating. *Research & Practice in Assessment, 8*(Summer), 13–26. Retrieved from <http://eds.b.ebscohost.com.library.esc.edu/eds/pdfviewer/pdfviewer?vid=1&sid=2c170537-bede-422a-b10f-3df29e48915e%40sessionmgr198&hid=112%5Cnhttp://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf>
- Mills, N., Pajares, F., & Herron, C. A. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning, 57*(3), 417–442. <https://doi.org/10.1111/j.1467-9922.2007.00421.x>
- Milton, J. (2006). Language lite? Learning French vocabulary in school. *French Language Studies, 16*, 187–205.
- Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: Comparing knowledge over time. *Journal of French Language Studies, 18*(3), 333–348. <https://doi.org/10.1017/S0959269508003487>
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, J., & Hopkins, N. (2005). *AuralLex*. Swansea: Swansea University.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review, 63*(1), 127–147. <https://doi.org/10.3138/cmlr.63.1.127>
- Milton, J., & Riordan, O. (2006). Level and script effects in the phonological and orthographic vocabulary size of Arabic and Farsi speakers. In P. Davidson, C. Coombe, D. Lloyd, & D. Palfreyman (Eds.), *Teaching and learning vocabulary in another language* (pp. 122–133). United Arab Emirates: TESOL Arabia.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. del M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.
- Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

- Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Psychological Science, 18*(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). The relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counselling Psychology, 38*, 30–38.
- Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research, 13*(1), 34–65. article. <https://doi.org/10.1191/026765897671676818>
- Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*, 287–311.
- Nation, K., Snowling, M., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *International Journal of Speech–Language Pathology, 9*, 131–139.
- Nation, P. (2000). Learning vocabulary in lexical sets: dangers and guidelines. *TESOL Journal, 1*–6.
- Nation, P. (2013). *Learning Vocabulary In Another Language*. Cambridge: Cambridge University Press.
- Nation, P. (2021). Thoughts on word families. *Studies in Second Language Acquisition, 43*, 969–972.
- Nicol, J. (2022). How fuzzy are L2 phonological representations? *Bilingualism, 25*(2), 230–231. <https://doi.org/10.1017/S1366728921000614>
- Nogueroles Lopez, M. (2017). Listening strategies instruction: effects on Hong Kong students' general strategic behaviour. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(6), 1–15. <https://doi.org/10.1186/s40862-017-0029-8>
- Ofqual. (2023). Provisional entries for GCSE, AS and A level: summer 2023 exam series. Retrieved November 24, 2023, from <https://www.gov.uk/government/statistics/provisional-entries-for-gcse-as-and-a-level-summer-2023-exam-series/provisional-entries-for-gcse-as-and-a-level-summer-2023-exam-series>
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26–45.
- Osberg, D., & Biesta, G. (2010). *Complexity theory and the politics of education*. Rotterdam ; Boston.
- Oxford, U. of. (2022). Management and protection of data collected for research purposes. Retrieved from

<https://researchsupport.admin.ox.ac.uk/files/bpg09datacollectionandmanagementpdf>

- Patterson, A. (2021). Predicting second language listening functor comprehension probability with usage-based and embodiment approaches. *International Journal of Bilingualism*, 25(3), 772–788.
- Patterson, A. (2022). The shape of a word: single word characteristics' effect on novice L2 listening comprehension. *Applied Linguistics Review*, 1–26. Retrieved from <https://doi.org/10.1515/applirev-2021-0157>
- Perfetti, C. A. (2017). Lexical quality revisited. In E. Segers & P. Van den Broek (Eds.), *Developmental perspectives in written language and literacy: In honor of Ludo Verhoeven* (pp. 51–67). Amsterdam: John Benjamins.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In P. Verhoeven, Ludo; Elbro, Carsten; Reitsma (Ed.), *Precursors of functional literacy*, (pp. 189–213). Amsterdam: John Benjamins. <https://doi.org/10.1075/swll.11.14per>
- Plonsky, L. (2023). Sampling and Generalizability in Lx Research: A Second-Order Synthesis. *Languages*, 8(1). <https://doi.org/10.3390/languages8010075>
- Prince, P. (2014). Listening comprehension: processing demands and assessment issues. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp. 93–108). Bristol: Multilingual Matters.
- Rahimirad, M., & Shams, M. R. (2014). The Effect of Activating Metacognitive Strategies on the Listening Performance and Metacognitive Awareness of EFL Students. *International Journal of Listening*, 28(3), 162–176. <https://doi.org/10.1080/10904018.2014.902315>
- Richards, J. C. (2005). Second Thoughts on Teaching Listening. *RELC Journal*, 36(1), 85–92. <https://doi.org/10.1177/0033688205053484>
- Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133–143). Abingdon: Routledge.
- Rose, H. (2019). Dismantling the Ivory Tower in TESOL: A Renewed Call for Teaching-Informed Research. *TESOL Quarterly*, 53(3), 895–905. <https://doi.org/10.1002/tesq.517>
- Rost, M. (2016). *Applied linguistics in action : teaching and researching listening*. Retrieved from https://books.google.fr/books/about/Teaching_and_Researching_Listening.html?id=ZpH4CgAAQBAJ&redir_esc=y
- Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (2023). Individual differences in L2 listening proficiency revisited: Roles of form, meaning, and use aspects of phonological vocabulary knowledge. *Studies in Second Language Acquisition*, 1–27. <https://doi.org/10.1017/S027226312300044X>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. (2001). Attention. In *Cognition and Second Language Instruction* (p. 1710362). Cambridge: Cambridge University Press. <https://doi.org/https://doi->

- Schmidt, R. (2010). Attention, Awareness, and Individual Differences in Language Learning. *Proceedings of CLaSIC 2010*, 4, 721–737.
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: an empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71–106.
- Sealey, A., & Carter, B. (2004). *Applied linguistics as social science*. London: Continuum.
- Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching*, X(3).
- Seyednozadi, Z., Pishghadam, R., & Pishghadam, M. (2021). Functional Role of the N400 and P600 in Language-Related ERP Studies with Respect to Semantic Anomalies: An Overview. *Archives of Neuropsychiatry*, 58(3), 249–252. <https://doi.org/10.29399/npa.27422>
- Sheppard, B., & Butler, B. (2017). Insights Into Student Listening From Paused Transcription. *The CATESOL Journal*, 29(2), 81–108.
- Shiu, L.-J., Yalçın, Ş., & Spada, N. (2018). Exploring second language learners' grammaticality judgment performance in relation to task design features. *System*, 72, 215–225. <https://doi.org/10.1016/j.system.2017.12.004>
- Simpson, K. (2017). *Can speech stream segmentation instruction improve listening comprehension and listening self-efficacy in lower intermediate learners?* Oxford.
- Simpson, K., & Rose, H. (2020). Complexity as a Valid Approach in 'Messy' Classroom Contexts: Promoting More Ecologically Rich Research on the Psychology of L2 Listening. In R. J. Sampson & R. S. Pinner (Eds.), *Complexity Perspectives on Researching Language Learner and Teacher Psychology* (pp. 136–151). Bristol: Multilingual Matters.
- Smit, N., Dijk, M. Van, Bot, K. De, & Lowie, W. (2022). The complex dynamics of adaptive teaching: observing teacher-student interaction in the language classroom, 60(1), 23–40.
- Smith, G. F. (2020). *An Investigation of vocabulary size, metacognition and individual differences in L2 listening comprehension*. University of Hawai'i at Manoa.
- Sprang, G., & Silman, M. (2013). Posttraumatic stress disorder in parents and youth after health-related disasters. *Disaster Medicine and Public Health Preparedness*, 7(1), 105–110. Retrieved from <https://doi.org/10.1017/dmp.2013.22>
- Stæhr, L. S. (2009). *Vocabulary Knowledge and Advanced Listening Comprehension in English As a Foreign Language*. *Studies in Second Language Acquisition* (Vol. 31). <https://doi.org/10.1017/S0272263109990039>
- Stubbs, B. (2017). Student performance analysis. Retrieved June 20, 2018, from <http://www.bstubbs.co.uk/gcse.htm>
- Tagashira, K., Kida, S., & Hoshino, Y. (2010). Hot or gelid? The influence of L1

- translation familiarity on the interference effects in foreign language vocabulary training. *Sytem*, 38(412–421).
- Taylor, W. A. (2000). Change-point analysis: A powerful new tool for detecting changes. Retrieved October 22, 2022, from <http://www.variation.com/cpa/tech/changepoint.html>
- Thompson, I., & Rubin, J. (1996). Can strategy instruction improve listening comprehension. *Foreign Language Annals*, 29(3), 331–342.
- Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). “Repeat as much as you can”: Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp. 143–163). Bristol: Multilingual Matters.
- Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge: Cambridge University Press.
- Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or Top-down Processing as a Discriminator of L2 Listening Performance. *Applied Linguistics*, 19(4), 432–451. <https://doi.org/10.1093/applin/19.4.432>
- Ushoida, E. (2021). Doing Complexity Research in the Language Classroom: A Commentary. In R. Sampson & R. Pinner (Eds.), *Complexity Perspectives on Researching Language Learner and Teacher Psychology* (pp. 269–282). Bristol: Multilingual Matters.
- Vafae, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, 42(2), 383–410. <https://doi.org/10.1017/S0272263119000676>
- van Dijk, M., Verspoor, M., & Lowie, W. (2011). Variability and DST. In *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 55–84). Amsterdam: John Benjamins Publishing Company.
- van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: an introduction. *Modern Language Journal*, 92(2), 179–199.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98(1), 3–53. <https://doi.org/10.1037//0033-295x.98.1.3>
- van Zeeland, H. (2014). Lexical inferencing in first and second language listening. *Modern Language Journal*, 98(4), 1006–1021.
- Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension. *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- Vandergrift, L. (1997). The Cinderella of Communication Strategies: Reception Strategies in Interactive Listening. *The Modern Language Journal*, 81(4), 494. <https://doi.org/10.2307/328892>
- Vandergrift, L. (2003). Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener. *Language Learning*, 53(3), 463–496. <https://doi.org/10.1111/1467-9922.00232>

- Vandergrift, L. (2004). Listening to Learn or Learning to Listen. *Annual Review of Applied Linguistics*, (24), 3–25.
- Vandergrift, L. (2015). Researching listening. In *Research methods in applied linguistics*. London: Bloomsbury.
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390–416. <https://doi.org/10.1111/lang.12105>
- Vandergrift, L., & Baker, S. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *The Canadian Modern Language Review*, 74(1), 79–100.
- Vandergrift, L., Goh, C., Mareschal, C., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, 56(3), 431–462.
- Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 Learners How to Listen Does Make a Difference: An Empirical Study. *Language Learning*, 60(2), 470–497. <https://doi.org/10.1111/j.1467-9922.2009.00559.x>
- Vanderplank, R. (2016). “Effects of” and “effects with” captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235–250. <https://doi.org/10.1017/S0261444813000207>
- Vanderplank, R., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. System (Vol. 40). New York: Routledge. Retrieved from <http://ezproxy.msu.edu/login?url=http://search.proquest.com/docview/1315888878?accountid=12598> http://za2uf4ps7f.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQ%3Allbashell&rft_val_fmt=info:ofi/fmt:kev:mtx:8
- Verspoor, M., & Bot, K. De. (2022). Measures of variability in transitional phases in second language development, 60(1), 85–101.
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, 92(2), 214–231. <https://doi.org/10.1111/j.1540-4781.2008.00715.x>
- Verspoor, M., Lowie, W., van Geert, P., van Dijk, M., & Schmid, M. S. (2011). Correlations and moving correlations. In *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 181–184). Amsterdam: John Benjamins.
- Verspoor, M., & van Dijk, M. (2011). Visualising Interactions Between Variables. In M. Verspoor, K. De Bot, & W. Lowie (Eds.), *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 85–98). Amsterdam: John Benjamins Publishing Company.
- Vogely, A. (1995). Performance of three authentic listening comprehension tasks. *The Modern Language Journal*, 79(1), 41–56.
- Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary

- knowledge and metacognitive awareness. *System*, 65, 139–150.
<https://doi.org/10.1016/j.system.2016.12.013>
- Wesche, M., & Paribakht, T. S. (2017). Introduction. In L. Ortega & Z. Han (Eds.), *Complexity Theory and Language Development* (pp. 1–10). Amsterdam: John Benjamins Publishing Company.
- Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, 41, 1056–1069.
- Wirtz, M. A., & Pfenninger, S. E. (2023). Variability and individual differences in L2 sociolinguistic evaluations: The GROUP, the INDIVIDUAL and the HOMOGENEOUS ENSEMBLE. *Studies in Second Language Acquisition*, 45(5), 1186–1209. <https://doi.org/10.1017/S0272263123000177>
- Wong, S. W. L., Leung, V. W. H., Tsui, J. K. Y., Dealey, J., & Cheung, A. (2021). Chinese ESL learners' perceptual errors of English connected speech: Insights into listening comprehension. *System*, 98, 1–10.
- Woore, R., Graham, S., Courtney, L., Porter, A., & Savory, C. (2018). *Foreign Language Education: Unlocking Reading (FLEUR)-A study into the teaching of reading to beginner learners of French in secondary school*.
- Wright, B., & Tennant, A. (1996). Sample size again. *Rasch Measurement Transactions*, 9(4), 468.
- Xu, F. (2011). Anxiety in EFL listening comprehension. *Theory and Practice in Language Studies*, 12(1).
- Yeldham, M. (2016a). Second Language Listening Instruction: Comparing a Strategies-Based Approach With an Interactive, Strategies/Bottom-Up Skills Approach. *TESOL Quarterly*, 50(2), 394–420. <https://doi.org/10.1002/tesq.233>
- Yeldham, M. (2016b). The decoding of word classes by L2 English listeners. *English Teaching and Learning*, 40(1), 49–78.
- Yeldham, M. (2017a). Developing a framework for investigating L2 listeners' longitudinal development. *International Review of Applied Linguistics in Language Teaching*, 0(0), 1–29. <https://doi.org/10.1515/iral-2017-0004>
- Yeldham, M. (2017b). Techniques for researching L2 listeners. *System*, 66, 13–26. <https://doi.org/10.1016/j.system.2017.03.001>
- Yeldham, M., & Gruba, P. (2016). The development of individual learners in an L2 listening strategies course. *Language Teaching Research*, 20(1), 9–34. <https://doi.org/10.1177/1362168814541723>
- Yu, H., & Lowie, W. (2020). Dynamic Paths of Complexity and Accuracy in Second Language Speech : A Longitudinal Case Study of Chinese Learners, (July 2019), 855–877. <https://doi.org/10.1093/applin/amz040>
- Zhang, W., & Yi Liao. (2023). The role of auditory processing in L2 vowel learning: evidence from recasts. *Humanities and Social Science Communications*, 502(10), 1–14.
- Zhang, X. (2013). Foreign language listening anxiety and listening performance:

- Conceptualizations and causal relationships. *System*, 41(1), 164–177.
<https://doi.org/10.1016/j.system.2013.01.004>
- Zhang, X. (2019). Foreign Language Anxiety and Foreign Language Performance: A Meta-Analysis. *Modern Language Journal*, 103(4), 763–781.
<https://doi.org/10.1111/modl.12590>
- Zhao, X., & Li, P. (2021). Fuzzy or Clear? A Computational Approach Towards Dynamic L2Lexical-Semantic Representation. *Frontiers in Communication*, 6(January), 1223–1238. <https://doi.org/10.3389/fcomm.2021.726443>
- Zheng, Q., & Gor, K. (2024). The Influence of Native Phonology, Allophony, and Phonotactics on Nonnative Lexical Encoding: A Vocabulary Training Study. *Language Learning*, 74(1), 146–183. <https://doi.org/10.1111/lang.12581>
- Zhou, S., & Rose, H. (2024). A longitudinal study on lecture listening difficulties and self-regulated learning strategies across different proficiency levels in EMI higher education. *Applied Linguistics Review*.
<https://doi.org/https://doi.org/10.1515/applirev-2023-0113>

15. Appendices

15.1. Piloting

15.1.1. Metacognitive Awareness Listening Questionnaire

The Metacognitive listening questionnaire (Vandergrift et al., 2006) was delivered as a subtest of all three full test pilots. It remained largely the same throughout the three iterations of the full test pilot, after it was ascertained that the participants understood both the instructions and the content of the questions. However, after the first pilot, for absolute clarity, it was decided to change the presentation somewhat: instead of presenting all statements on a grid, with the six-point Likert scale on the x axis and the statements on the y-axis, each statement was presented individually, after which participants had to choose which point on the Likert scale was the most appropriate. The font and font size were altered to make it more readable, and the question order was randomised to ensure that participants did not copy one another.

The statements were as follows:

Type scale	Strategy or belief/perception
Person Knowledge	I find that listening in French is more difficult than reading, speaking, or writing in French.
Person Knowledge	I feel that listening comprehension in French is a challenge for me.
Person Knowledge	I don't feel nervous when I listen to French.
Directed Attention	I focus harder on the text when I have trouble understanding.
Directed Attention	When my mind wanders, I recover my concentration right away.
Directed Attention	I try to get back on track when I lose concentration.
Directed Attention	When I have difficulty understanding what I hear, I give up and stop listening.
Mental Translation	I translate in my head as I listen.
Mental Translation	I translate key words as I listen.
Mental Translation	I translate word by word, as I listen.
Problem-Solving	I use the words I understand to guess the meaning of the words I don't understand.

Problem-Solving	As I listen, I compare what I understand with what I know about the topic.
Problem-Solving	I use my experience and knowledge to help me understand.
Problem-Solving	As I listen, I quickly adjust my interpretation if I realize that it is not correct.
Problem-Solving	I use the general idea of the text to help me guess the meaning of the words that I don't understand.
Problem-Solving	When I guess the meaning of a word, I think back to everything else that I have heard, to see if my guess makes sense.

15.1.2. Additional question

The only alteration made was after the open question ‘has anything happened today that would have affected your performance on this test’; initially, this contained the additional wording ‘if yes, say what’, with an input box into which participants could type an answer. This was changed to a simple yes / no answer. If participants answered no, they were moved straight onto the next question, whereas only if they answered yes were they then given an open input box in which they could state what the problem might be.

15.1.3. Working memory tests

I initially wanted to include a working memory test and created a reverse digit span test to include in the test battery. Participants would listen to series of numbers and then enter them in reverse into the dialogue box within the Qualtrics test battery. First they would hear two series of three numbers, then two of four four, etc, up until two series of eight. After each series the dialogue box would appear. My plan had been to then award a score between 3 and 8, depending on the highest number that they reliably reversed.

However, it was clear that by introducing the typing (written) element to this – in order to collect the data en masse from 80 participants – I introduced confounds of echoic memory and strategy into this. In piloting I watched many participants use their echoic memory to type out the numbers in the order in which they heard them, then use the ‘cut’ and ‘paste’ keys to reverse the order. Combined with the fact that I felt I was already

collecting too much data, I then chose to remove the Working Memory tests from my model.

15.2. Which paused translation test when

(where the number refers to the time point and the first, second or third test – so 18.3 means it was the third listening test at time 18)

1. Access Studio track 8 shortened 1.3 2.1
2. Access Studio track 7 1.1 4.2
3. Access Studio 4 1.2
4. Studio M2 3B 2.2, 3.2 6.2
5. St2M1-2A 8.1
6. G78 3.1
7. G32 6.1
8. F9 4.1, 7.2, 10.2
9. F46 7.1
10. F28 5.1
11. F25 2.3 5.2 9.1
12. E65 9.2
13. E56 6.3 13.1 19.1
14. E3 5.3 8.2 11.1
15. E26 4.3 7.3 12.1 15.1
16. E111 3.3 10.1 14.1 18.2
17. D8 8.3 12.2 16.2 20.2
18. D47 11.2, 17.1
19. D31 14.2
20. D3 10.3 15.2 19.2
21. D25, 9.3 13.2 18.1
22. C55 13.3
23. C48 11.3 14.3 17.2
24. C41 16.3
25. C2 15.3
26. C17 12.3
27. C11 17.3
28. Studio M3 assessment pack 18.3
29. Studio 3 Assessment pack Module 5 test 1A 19.3
30. Studio 3 St342A

15.3. Full transcripts of all listening tests, plus English translations

(where the number refers to the time point and the first, second or third test – so 18.3 means it was the third listening test at Time 18)

Access Studio track 8 1.3 2.1

Salut, je m'appelle Amélie. / Moi, j'adore mon professeur Madame Foulon. / J'aime les chiens / mais je n'aime pas les poissons rouges! / Je déteste les jeux vidéo, / c'est nul, / mais j'adore la gymnastique, / c'est super. Comme couleur, je préfère le rouge / et je déteste le marron.

Hello, my name is Amélie / I love my teacher Madame Foulon / I like dogs / but I don't like goldfish. / I hate video games / they're rubbish / but I love gymnastics, / it's super / As a colour, I prefer red / and I hate brown.

Access Studio track 7

J'adore le skate et j'aime aussi le rugby. / Moi, j'aime le tennis et j'adore la danse. / J'adore le vélo et aussi le foot. / Moi, j'aime la gymnastique et j'adore la danse. / J'aime le judo et les jeux vidéo. / Moi, j'adore le vélo et la musique. / J'adore le rugby et les jeux vidéo.

I love skateboarding and I also love rugby. / Me, I like tennis and I love dance / I love cycling and also football / Me, I love gymnastics and I love dance / I like judo and video games / I love cycling and music / I love rugby and video games.

Access Studio 4 1.2

– Comment t'appelles-tu ? Je m'appelle Amira. / Quel âge as-tu ? J'ai 11 ans. / C'est quand, ton anniversaire ? C'est le 15 décembre.

– Comment t'appelles-tu ? Je m'appelle Pierre. / Quel âge as-tu ? J'ai 14 ans. / C'est quand, ton anniversaire ? C'est le 10 février.

– Comment t'appelles-tu ? Je m'appelle Rose. / Quel âge as-tu ? J'ai 19 ans. / C'est quand, ton anniversaire ? C'est le premier Avril.

What's your name ? My name is Amira / How old are you? I'm 11. When's your birthday? It's the 15th December.

What's your name. My name is Pierre / How old are you? I am 14 / When's your birthday? It's 10th February.

What's your name? My name is Rose. / How old are you? I'm 19. When's your birthday? It's 1st April.

Studio M2 3B 2.2, 3.2 6.2

J'aime bien le collège et mes matières / mais je n'aime pas les sciences. / C'est difficile. / Les profs sont gentils. / J'aime beaucoup mon prof de maths. / Le matin, on commence à huit heures. / La récréation est à dix heures trente.

I like school and my subjects / but I don't like science / it's difficult. / The teachers are nice / I like my maths teacher a lot / In the morning, we start at 8am / Breaktime is at 10.30.

F25 2.3 5.2 9.1

Je joue du piano / Moi j'aime jouer de la guitare / Je ne joue pas d'un instrument / Mais je suis chanteur dans un groupe / J'adore écouter de la musique classique / Je préfère télécharger des chansons.

I play piano / Me I like playing the guitar / I don't play an instrument / But I am a singer in a band / I love listening to classical music / I prefer downloading songs.

G78 3.1

Je m'appelle Paul Poirot / ça s'écrit P O I R O T / J'ai seize ans / au collège je préfère le dessin / mon sport préféré c'est le volley / J'aime aller au cinéma

My name is Paul Poirot / That's spelled P O I R O T / I'm 16 / at school I prefer art / my favourite sport is volleyball / I like going to the cinema.

F9 4.1, 7.2, 10.2

Je joue souvent au football / moi je préfère la natation / j'adore mon cheval je fais souvent de l'équitation / j'aime aller au bord de la mer pour faire de la voile / en été je pratique le ski nautique.

I often play football / me I prefer swimming / I love my horse and I often go horseriding / I like going to the seaside to do sailing / In the summer I do waterskiing.

E26 4.3 7.3 12.1 15.1

Je prends des photos avec mon portable. / Je préfère envoyer souvent des textos. / Moi, je sauvegarde de la musique sur mon portable. / Je bavarde avec mes amis.

I take photos with my mobile / I prefer often sending texts / Me, I save music on my phone / I chat with my friends.

E111 3.3 10.1 14.1 18.2

Marc, ça va au collègue? / Tu aimes ça? / L'histoire, c'est ma matière préférée / parce que c'est utile, / mais je ne supporte pas l'anglais. / Moi, j'attends toujours la récré./ car je vais à la bibliothèque tous les jours.

Mark, how is school ? / Do you like it ? / History is my favourite subject / because it's useful / but I can't stand English. / Me, I'm always looking forward to breaktime / because I go to the library every day.

F28 5.1

Le matin, j'adore manger des céréales. / Je n'aime pas le chou-fleur. / Je mange beaucoup de fromage. J'aime ça. / Les chips? Moi j'adore les chips! / Le poisson? Non, je ne mange jamais de poisson. Beurk!

In the morning I like eating cereal / I don't like cauliflower / I eat a lot of cheese. I love that / Crisps? I love crisps! / Fish? No, I never eat fish! Yuck!

E3 5.3 8.2 11.1

Mes frères et sœurs sont formidables / mais quelquefois ma grande sœur, Laura, est méchante. / Quand je regarde la télévision elle parle tout le temps. / Ma petite sœur, Mimi, est adorable. / J'aime jouer avec elle. / Quand je joue au football avec mon frère Mickaël, ça va bien, / mais quelquefois il joue seulement avec ses copains. / Ma sœur préférée, c'est ma sœur aînée, Céline.

My brothers and sisters are great / but sometimes my big sister, Laura, is mean / When I watch TV she talks all the time / My little sister Mimi is adorable / I like playing with her / When I play football with my brother Mickael, it's fine / but sometimes he only plays with his friends / My favourite sister is my oldest sister, Céline.

G32 6.1

Je vais dans de grandes villes / Je vais à la plage / On va à la montagne / Et nous, à la campagne / Moi, je vais dans un petit village.

I go to big cities / I go to the beach / We go to the mountains / and us, to the countryside / Me, I go to a small village.

E56 6.3 13.1 19.1

J'habite dans une station de ski à la montagne. / J'habite dans une petite ville / dans l'est de la France avec ma famille. / Moi, j'habite avec ma mère. / Notre ville est grande et industrielle et je ne l'aime pas. / J'habite dans un port de pêche dans le sud de la France. / J'habite dans une petite ville au bord de la mer / avec mon frère et mes parents.

I live in a ski resort in the mountains / I live in a small town / in the east of France with my family / Me, I live with my mum / Our town is big and industrial and I don't like it / I live in a fishing village in the south of France / I live in a little town by the sea / with my brother and my parents.

F46 7.1

Mon ami Paul habite en Espagne. / Ma cousine Lily habite au pays de Galles avec sa famille. / Mon oncle Georges habite aux États-Unis / dans une très grande maison / et il aime le temps chaud là-bas / Ma grand-mère habite en Angleterre / dans un petit appartement chic.

My friend Paul lives in Spain / My cousin Lily lives in Wales with her family / My uncle George lives in the USA / in a very big house / and he likes the hot weather over there / My grandmother lives in England / in a little chic apartment.

D8 8.3 12.2 16.2 20.2

J'apprends à jouer de la guitare. / d'habitude je reste à la maison / et je regarde un film d'épouvante avec mes frères. / J'adore ça. / j'aime faire une promenade avec mon chien. / Mon passe-temps préféré, c'est écouter de la musique rap dans ma chambre. / je fais du vélo avec mes copines.

I'm learning to play the guitar / Usually I stay at home / and I watch a horror film with my brothers / I love that / I like to go for a walk with my dog / My favourite hobby is listening to rap music in my room / I go cycling with my friends.

St2M1-2A 8.1

Je regarde beaucoup la télé. / J'adore les émissions musicales: / elles sont super! / Puis j'aime aussi les documentaires, / surtout sur les animaux. / Avec ma mère, je regarde souvent les jeux télévisés. / Ils sont stupides mais très amusants, / j'adore ça! / Je ne regarde jamais les émissions de sport. C'est barbant!

I watch TV a lot / I love music programmes, they're great / Then I also like documentaries / especially about animals. / With my mum, I often watch gameshows / They are stupid but very funny / I love that / I never watch sports programmes / They're boring.

E65 9.2

Je n'utilise pas les sacs en plastique. / Je vais au collège à pied tous les jours. / Je prends toujours une douche. / Je recycle mes magazines. / Je vais voyager en vélo.

I never use plastic bags / I go to school on foot every day / I always take a shower / I recycle my magazines / I'm going to travel by bike.

D25, 9.3 13.2 18.1

Aline, tu vas regarder la télé ce soir? / Non, j'aime regarder les films français, / mais ce soir il y a beaucoup de feuilletons américains. / Et toi, Frédéric? / il y a deux bons documentaires sur l'environnement. / mais l'environnement ne m'intéresse pas. / il y a des dessins animés, mais c'est beaucoup trop tard. / Bon, alors moi, je vais regarder la télé / Et toi, tu vas lire.

Aline, are you going to watch TV this evening / No, I like watching French films / but this evening there are a lot of American series / And you, Frederic? There are two good documentaries about the environment / But I'm not interested in the environment / There are cartoons / but it's a lot too late / So me, I'm going to watch TV / And you, you are going to read.

D3 10.3 15.2 19.2

J'adore ma famille. / Tout le monde est formidable. / Nathan, le plus petit de la famille, est très amusant. / Sa sœur aînée, Martine, est très travailleuse. / Ma fille cadette, Sylvie, adore s'amuser avec ses amies. / Et mon mari, Jacques? / C'est un homme très gentil /et pas du tout égoïste.

I love my family / Everyone is great / Nathan, the youngest in the family, is very funny / His big sister Martine is very hardworking / My older daughter, Sylvie, loves having fun with her friends / And my husband Jack? / He's a very kind man / and not at all selfish.

D47 11.2, 17.1

J'habite dans une ferme à la campagne. / Moi, j'habite dans une maison individuelle. / Nous avons un jardin, mais il n'y a pas de garage. / J'habite dans un petit appartement au rez-de-chaussée / Moi, j'habite une maison jumelée. / Notre garage est grand, mais nous n'avons pas de jardin. / Moi aussi, j'habite une maison jumelée, / mais nous n'avons pas de garage et pas de jardin.

I live on a farm in the countryside / Me, I live in a detached house / We have a garden, but there isn't a garage / I live in a small flat on the ground floor / Me I live in a semi-detached house / Our garage is big, but we don't have a garden / Me too, I live in a semi-detached house, but we don't have a garage and we don't have a garden.

C48 11.3 14.3 17.2

Il y a une grande armoire dans ma chambre / et j'aime ça /car il y a plein d'espace pour tous mes vêtements, /Je dois partager ma chambre avec mon frère. / c'est bien car on peut bavarder tard le soir, / Notre chambre est au quatrième étage, / alors on peut jouer la musique très forte sans problèmes, / mais, la salle de bains est loin de la chambre.

There is a big wardrobe in my room / and I like that / because there is plenty of space for all my clothes / I have to share my room with my brother / it's good because we can talk late into the evening / Our room is on the fourth floor / so we can play music very loud with no problem / but the bathroom is far from the bedroom.

C17 12.3

Aline, que manges-tu au petit déjeuner? / Moi, je mange une tartine, / mais j'adore mettre du miel sur ma tartine. / Et toi Pierre? / je bois du thé avec du lait, / mais le weekend je préfère prendre deux œufs /avec du pain grillé. / – Et toi Bernadette? / – Je ne mange rien avant l'école. / Je mets un sandwich au jambon dans mon sac / et je pars.
Aline, what do you eat for breakfast? / I eat bread and butter / but I love putting honey on my bread / and you Pierre? / I drink tea with milk / but at the weekend I prefer having two eggs / with some toast / And your Bernadette? / I eat nothing before school / I put a ham sandwich in my bag / and I leave.

C55 13.3

Mon frère Dominique ne cesse jamais de parler / mais je m'entends bien avec lui. / Mon copain Mohammed me fait toujours rire / et il est vraiment rigolo. / Mon ami Yvon est toujours actif. / À la maison, il aide ses parents / et il a beaucoup d'énergie.

My brother Dominic never stops talking / but I get on well with him / My friend Mohamed always makes me laugh / and he is really funny / My friend Yvon is always active / At home, he helps his friends / and he has a lot of energy.

D31 14.2

Kevin, où est-ce que tu passes tes vacances? / Ma famille a un petit appartement au centre-ville à Paris. / nous allons toujours dans un grand hôtel en Espagne. / J'adore aller chez ma cousine. / Elle n'habite pas loin de Bordeaux. / Je pars avec mes amis. / On va toujours à une auberge de jeunesse. / On n'a pas beaucoup d'argent, tu sais. / Ma famille loue une maison à la campagne.

Kevin, where do you spend your holidays? / My family has a little flat in the centre of Paris / We always go to a big hotel in Spain / I love going to my cousin's / She lives not far from Bordeaux / I go away with my friends / We always go to a youth hostel / We don't have a lot of money, you know / My family hires a house in the countryside.

C2 15.3

Mon frère vient d'acheter une souris. / Je la trouve mignonne / Quand j'avais onze ans ma mère m'a acheté un cochon d'Inde / Ma petite sœur n'est pas contente. / Elle voudrait avoir un chien mais / moi je ne les aime pas du tout. / J'aimerais avoir un cheval car j'adore l'équitation, / Moi, j'ai un poisson rouge mais je le trouve barbant / car il n'est pas actif.

My brother has just bought a mouse / I find it sweet / when I was 11 my mother bought me a guinea pig / My little sister is not happy / She would like to have a dog / but I don't like them at all / I would like to have a horse because I love riding / Me, I have a goldfish, but I find it boring / because he is not at all active.

C41 16.3

J'ai faim! / Bon, on peut aller dans cette boulangerie acheter des gâteaux. / Beurk! Je déteste les choses sucrées. / Tu es vraiment difficile. / Là, on vend des frites et des sandwiches. / Beurk! Je déteste les sandwiches. / Mais tu adores les frites. / Oui, c'est vrai. / Tu veux une saucisse avec des frites? / Pas spécialement. / mais je voudrais bien prendre une bouteille d'eau.

I'm hungry / OK, we can go to that bakery and buy cakes / Yuck. I hate sweet things / you are really difficult / There they sell chips and sandwiches / Yuck, I hate sandwiches / But you love chips / Yes, that's true / Do you want a sausage with your chips? Not especially / But I would like a bottle of water.

C11 17.3

J'ai trouvé la glace à la vanille super. / À mon avis, le poulet rôti était froid, / alors je n'irai jamais dans ce café à l'avenir. / Les frites m'ont beaucoup plu / car elles n'étaient pas trop petites. / Pour moi, le gâteau au chocolat était bien / mais le prix était trop cher. / Selon moi, le sandwich au bœuf était vraiment savoureux.

I found the vanilla ice-cream super / In my opinion, the roast chicken was cold / so I will never go to that café in the future / I really liked the chips / because they were not too small / For me the chocolate cake was good / but the price was too expensive / According to me, the beef sandwich was really tasty.

Studio M3 assessment pack 18.3

Moi, cette année, je pars en vacances avec ma famille en Espagne. / On ira à la campagne. / J'aime bien la campagne. / Mon frère aime faire du VTT, / mais moi, je

préfère faire des randonnées dans la forêt. / On fera du camping. / L'année dernière, je suis parti en vacances avec mes amis. / On est allés au bord de la mer. / C'était génial et je me suis bien amusé.

Me, this year I am going on holiday with my family to Spain / We're going to the countryside / I like the countryside a lot / My brother likes to do mountain biking / but me, I prefer to go for walks in the forest / We will go camping / Last year, I went on holiday with my friends / we went to the seaside / It was great, and I had a lovely time.

Studio 3 Assessment pack Module 5 test 1A 19.3

Bien sûr, j'ai un portable. / J'ai aussi un ordinateur dans ma chambre / et je peux l'utiliser jusqu'à neuf heures du soir. / Par contre, je n'ai pas le droit de regarder la télé dans ma chambre. / Je peux sortir avec mes amis le weekend, / mais je ne peux pas rentrer seule. / Je n'ai pas le droit de sortir en semaine. / J'ai le droit de manger au MacDo le samedi avec mes amis. / J'ai aussi le droit de jouer à des jeux vidéo avec mon frère.

Of course I have a mobile phone / I also have a computer in my room / and I can use it until nine in the evening / On the other hand, I'm not allowed to watch TV in my room / I can go out with my friends on the weekend / but I'm not allowed to come home alone / I am not allowed to go out in the week / I'm allowed to eat at McDonald's on Saturdays with my friends / I'm also allowed to play video games with my friends.

St342A 20.3

Alors, William, que feras-tu cet été? // J'irai au bord de la mer / et je ferai du ski nautique. Mes parents aiment faire de l'escalade, / mais cette année, on fera tous un stage de voile. / Ce sera très fatigant / mais amusant, j'espère. / Normalement, on fait des sports extrêmes, / mais en juillet, / j'irai voir les animaux sauvages / dans un grand parc en Afrique. / Moi, j'irai sur une île tropicale. / Mon copain descendra l'Amazone en canoë, / mais moi, je préfère le calme et le repos!

So William, what will you do this summer? I'll go to the seaside / and I'll do waterskiing. My parents like climbing / but this year, we're all going to do a sailing course / It will be very tiring / but fun, I hope. / Normally, we do extreme sports / but in July / I will go to see wild animals / in a big park in Africa / Me, I will go to a tropical island / My friend will go down the Amazon in a canoe / but me, I prefer calm and rest.

15.4. Full list of AuralLex tests

Test 3

1. bière	21. soie	jupaire
2. inquiet	22. à	jivercle
3. résultat	23. Roti	métracte
4. Gâcher	24. voyou	morler
5. vraiment	25. Client	nolois
6. voyager	26. Docteur	octe
7. charcuterie	27. raide	plassard
8. milieu	28. hors d'œuvre	préjate
9. côte	29. rembourser	
10. Numéro	30. maladie	
11. réussir	31. paraître	
12. normale	32. Bowling	
13. anglais	33. librairie	
14. Bleu	34. Mathématique	
15. ombre	35. Heureux	
16. Camion	36. meilleur	
17. descendre	37. licence	
18. Repos	38. passe-temps	
19. ensoleillé	39. Chrétien	
20. Casque	40. Palais	

Test 5

1. croisière	21. Pessimiste	sortape
2. Industriel	22. complet	tellene
3. scolaire	23. dormir	tinquant
4. crêpe	24. Conférence	triparoix
5. rez-de-chaussée	25. Tirer	valline
6. Saignant	26. fumer	veindre
7. lundi	27. occupé	dupoux
8. inquiet	28. boucher	cutat
9. Cause	29. gaz	
10. Licencier	30. Copie	
11. Toussaint	31. Caractère	
12. Couramment	32. pressé	
13. déjeuner	33. coûter	
14. Juif	34. Humilier	
15. grosse	35. incroyable	
16. Salade	36. suite	
17. bague	37. musculation	
18. Icône	38. reconnaissant	
19. Cahier	39. Privé	
20. Exprès	40. agence	

Test 6

1. Distributeur	21. Peur	tecler
2. soigner	22. escargot	méot
3. pauvre	23. Huitre	bonin
4. truite	24. escalier	pallier
5. court	25. fils	ferrand
6. poulet	26. internet	trépanier
7. robe	27. humour	josserand
8. soin	28. égoïste	bour
9. Dessiner	29. Poche	
10. lumière	30. Copain	
11. tourisme	31. Baguette	
12. désolé	32. Lui	
13. vide	33. bouclé	
14. vrai	34. beaucoup	
15. me	35. supporter	
16. Alpes	36. jamais	
17. nom	37. cheval	
18. étudier	38. dessinateur	
19. pantalon	39. Piéton	
20. auto	40. Édimbourg	

Test 7

1. grand	1. aventure	catteau
2. Déchirer	2. coiffeur	abadie
3. assiette	3. nom	vignaud
4. location	4. habile	gougeon
5. mépriser	5. forme	dalle
6. normale	6. Péage	tixier
7. respirer	7. quoi	scorailles
8. prix	8. Fixe	asselin
9. accro	9. attention	
10. Mars	10. bonbon	
11. agence	11. ainsi	
12. relaxer	12. cerise	
13. Café	13. plat	
14. trajet	14. Professionnel	
15. dépenser	15. Racaille	
16. beurre	16. barbe	
17. où	17. surface	
18. manquer	18. marrant	
19. chanteur	19. bruit	
20. randonnée	20. vendredi	

15.5. Full list of AuraGram tests

Test A

		Grammatical	Ungrammatical	Don't know
1.	Le garçon est curieux.			
2.	La rouge pomme est délicieuse.			
3.	Ma mère est plus sympa que mon père.			
4.	Ma frère est stupide.			
5.	Tu joues bien.			
6.	Demain je joué au foot.			
7.	C'est très intéressant.			
8.	Ils mangeons une banane.			
9.	Nous nous levons à onze heures			
10.	La musique que est super.			
11.	Le prof que j'aime.			
12.	J'aime toi.			
13.	Elle n'aime pas le foot.			
14.	Nous adorons joue au foot.			
15.	Il voyage en bus.			
16.	J'ai mange une banane.			
17.	Je vais aller en France.			
18.	Je suis joue au foot.			
19.	Il est mon ami depuis deux semaines.			
20.	Il faut que je vais en France.			

Test C

		Grammatical	Ungrammatical	Don't know
1.	La fille est curieuse.			
2.	Le français film est intéressant.			
3.	Je suis plus intelligent que toi.			
4.	Ma école est super.			
5.	Je mange lentement.			
6.	Hier je vais manger une pomme.			
7.	J'ai beaucoup de devoirs.			
8.	Nous regarde la télé.			
9.	Elle s'appelle Marie.			
10.	La banane que est jaune.			
11.	La télévision que je regarde.			
12.	Je mange il.			
13.	Nous ne pensons pas ça.			
14.	Elle adore va en ville.			
15.	Nous jouons au rugby.			
16.	Il a allé en France.			
17.	Ils vont regarder un film.			
18.	Je suis regarder la télévision.			
19.	Je parle français depuis deux ans.			
20.	Il faut que je fais mes devoirs.			

Test E

		Grammatical	Ungrammatical	Don't know
1.	Mon père est intelligent.			
2.	La américaine musique est super.			
3.	Mon ami est plus riche que moi.			
4.	Mon parents sont sympa.			
5.	Il danse rapidement.			
6.	Hier je regarde la télé.			
7.	C'est assez cool.			
8.	Elles être belles.			
9.	Je m'appelle Bob.			
10.	Le chat que est noir.			
11.	La musique que j'écoute.			
12.	Il regarde moi.			
13.	Ils ne sont pas stupides.			
14.	Ils aiment danse.			
15.	Je regarde un film.			
16.	J'ai resté à la maison.			
17.	Il va manger au restaurant.			
18.	Je suis fais mes devoirs.			
19.	Je joue du violon depuis 3 ans.			
20.	Il fait que je vais aux toilettes.			

Test F

		Grammatical	Ungrammatical	Don't know
1.	La fille est curieux.			
2.	La radio anglaise est stupide.			
3.	Je suis intelligenter que toi.			
4.	Mon école est super.			
5.	Elle parle douce.			
6.	Demain, je vais manger une pomme.			
7.	J'ai beaucoup des devoirs.			
8.	Nous regardons la télé.			
9.	Il est appelle David.			
10.	La banane qui est jaune.			
11.	Le prof qui j'aime.			
12.	Je l'aime.			
13.	Je non suis timide.			
14.	Il préfère regarder la télévision.			
15.	Ils faisons de la danse.			
16.	Il est allé en France.			
17.	Ils vont regarde un film.			
18.	Je vais au collège.			
19.	Il a parlé depuis cinq minutes.			
20.	Il faut que tu sois sage.			

Test G

		Grammatical	Ungrammatical	Don't know
1.	Ma sœur est grande.			
2.	La jaune banane est horrible.			
3.	Tu es moins généreux que moi.			
4.	Mon mère est jolie.			
5.	Je travaille calmement.			
6.	Hier, je vais aller à l'école.			
7.	Il fait trop chaud.			
8.	Je j'aime le foot.			
9.	Je me lave avant l'école.			
10.	La fille que est curieuse.			
11.	La couleur que je préfère.			
12.	Nous regardons est.			
13.	Tu n'es pas français.			
14.	J'aime joue de la guitare.			
15.	Je vais en ville.			
16.	Ils avons visité Paris.			
17.	Nous allons jouer au rugby.			
18.	Je suis mangeant mon diner.			
19.	Je fais du ballet depuis cinq ans.			
20.	Il faut que tu prends ton diner.			

Test I

		Grammatical	Ungrammatical	Don't know
1.	Ma mère est intelligente.			
2.	La anglaise radio est stupide.			
3.	Mon prof est le plus stricte.			
4.	Mon amis sont supers.			
5.	Elle parle doucement.			
6.	Demain, j'ai visité Paris.			
7.	J'ai un peu de chocolat.			
8.	Ils mangent une banane.			
9.	Il s'appelle David.			
10.	Le prof que est stricte.			
11.	La leçon que je déteste.			
12.	J'aime c'est.			
13.	Je ne suis pas timide.			
14.	Il préfère regarde la télévision.			
15.	Ils font de la danse.			
16.	Nous vu le film.			
17.	Nous allons parler en français.			
18.	Je suis vais au collège.			
19.	Il parle depuis cinq minutes.			
20.	Il faut que tu es sage.			

Test J

		Grammatical	Ungrammatical	Don't know
1.	Ma mère est intelligent.			
2.	La pomme rouge est délicieuse.			
3.	Tu es généreuser que moi.			
4.	Mes parents sont sympa.			
5.	Je travaille calme.			
6.	Hier, j'ai visité Paris.			
7.	C'est un très intéressant.			
8.	On aimons le cinéma.			
9.	Elle est appelle Marie.			
10.	Le chat qui est noir.			
11.	La télévision qui je regarde.			
12.	Nous la regardons.			
13.	Tu es ne français pas.			
14.	Elle adore aller en ville.			
15.	Je vas en ville.			
16.	Ils ont visité Paris.			
17.	Il vais manger au restaurant.			
18.	Je fais mes devoirs.			
19.	Il a été mon ami depuis deux semaines.			
20.	Il faut que tu prennes ton diner.			

15.6. Excerpts from reflexive diary for coding the paused translation

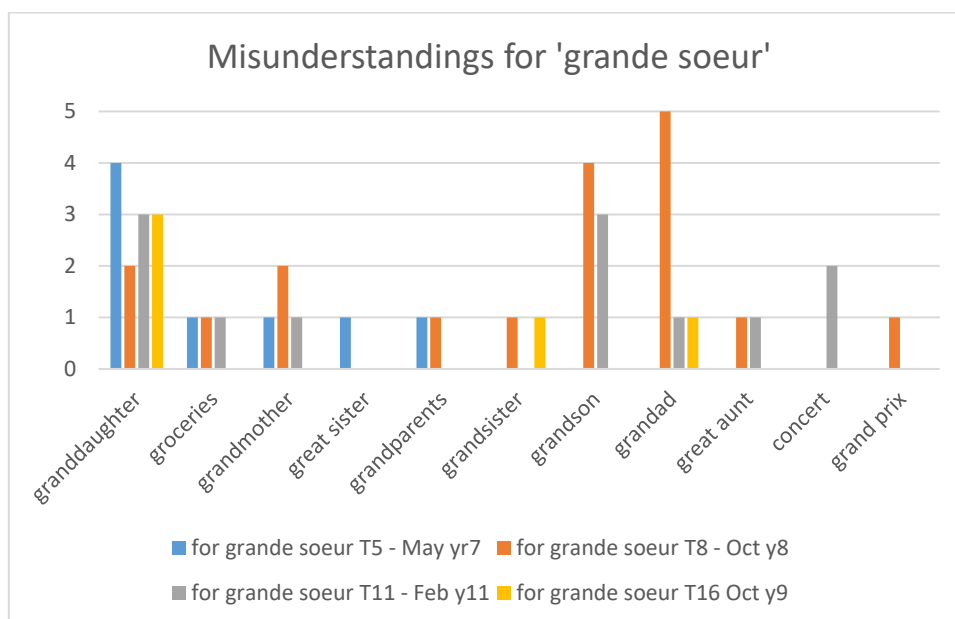
[copied and pasted from my notes]

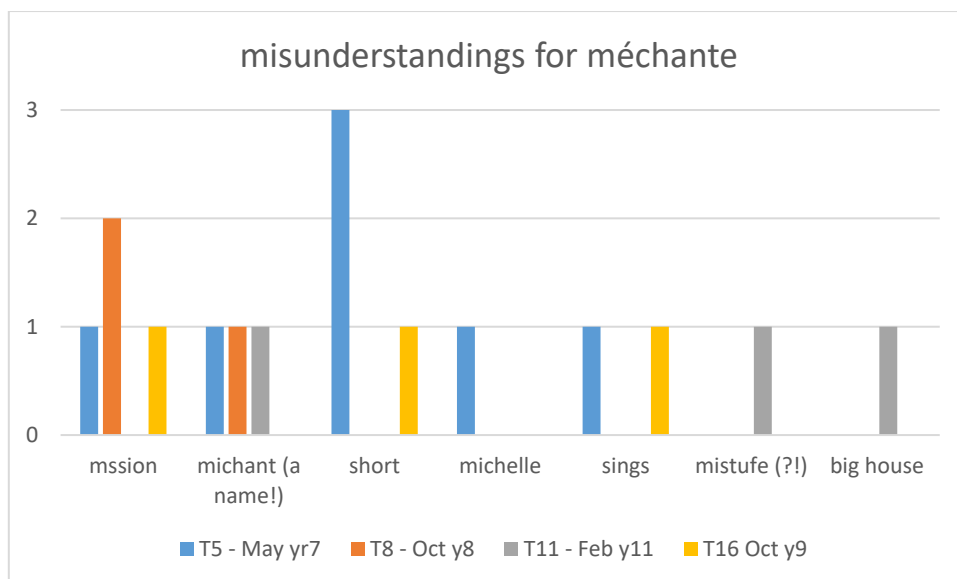
E3: *Mes frères et sœurs sont formidables...* lots of failure to detect plural at T5 – went through all four to check for the obvious ones first (e.g. didn't detect plural, missed end)

Mais quelquefois ma grande sœur Laura est méchante – function words missing – there are HEAPS of interesting things in this one. Crucially students really struggling with 'ma grande sœur' and defaulting to a family member with 'grand' in the word e.g. *grand-mère / grand-père*.

Then did 'recognised single word / false friend (for granddaughter etc)' only

Méchante caused a lot of problems. Translations included...





Quand je regarde la télé, elle parle tout le temps. Started with initial function word missing (lots of this all the way through! Also – random other verbs for ‘like’. Muddling of time words – this became a lot more frequent at later time points – probably at earlier time points they just didn’t notice them at all. Then just random sentences made with the idea of telly in them.

Ma petite sœur Mimi est adorable – start with ‘missed the end’. Some with content word missing when just ‘*adorable*’ is missing. Vocab issues with ‘*adorable*’ – horrible, terrible etc. Various people imposing ‘tablets’. Wrong family members. Pizzas, teachers, sir...

J’aime jouer avec elle – difficulties with ‘*elle*’ (seen as a name – Elle / Raquel) – some missing ‘I like’

Quand je joue au foot avec mon frère Mickael, ça va bien

Mais quelquefois il joue seulement avec ses copains – there’s a thing about word boundaries here – someone has come up ‘soleil’ for ‘seulement’ where it really identifies the relationship between word boundaries and vocab knowledge – [struggling to put a finger on exactly what, just now...] T11 – six incidences of Solomon and one of ‘Sully’ (T8 – no Solomons)

Ma sœur préférée, c'est ma sœur aînée, Cécile – question is – when they translate ‘*ma sœur préférée*’ as ‘my sister prefers’ is that a grammar mistake or a vocab? – it means they’re ignoring the ‘c’est’ to bend the meaning to their understanding. I’ve coded it under ‘favourite-prefer in vocab’ but I suppose I can go back over it from there... wonder about adding it to grammar though. Possibly later...

E56

J'habite dans une station de ski à la montagne – lots of issues not knowing ‘station’ for ski resort.

J'habite dans une petite ville – lots of ‘village’ – a few ‘villa’ or house.

Dans l'est de la France avec ma famille – mainly missing out ‘*l'est*’ altogether. This one also lots of ‘missed the start, got the end’ – suggests maybe it’s just about salience – as Robert said ‘transfer of training’ – they hear ‘*avec ma famille*’ so much in so many contexts, that it has become salient, and it sticks well. Really suggests they need A LOT more input before things make sense, for most students at least. Here there is also an increase in wrong tense – I wonder whether when the whole idea of ‘France with my family’ lends itself to a past tense – students transferring their own experience (I went to France with my family as a holiday type thing) more so than imagining that here is a person saying they live in France with their family, which is less relatable.

Moi j'habite avec ma mère. One or two beaches / seashores. One or two ‘brothers’ (wonder whether there should be a category for ‘wrong family member?’)

Notre ville est grande et industrielle et je ne l'aime pas – at T6 and T13, plenty of ‘Australia’ for ‘*industrielle*’. Not seen at all at T19. Question of maturity with ‘big words’ like ‘*industrielle*’. Still lots of villages and villas / houses for ville.

J'habite dans un port de pêche dans le sud de la France – interesting in T6 that there's lots of nods to maritime activities – sailing, going to the sea – without quite getting it.

These disappear. Peaches come up, inevitably. Occasional nuttury (I live in a pear)

J'habite dans une petite ville au bord de la mer. When students write 'I live in a small town with my mother' it's clear evidence that content words are more salient than function words. The '*mer*' overrides the lack of '*avec*' – particularly as there's plenty of evidence with most listeners that they do know *avec*.

Thought – the 'wrong tense' seems very random – see whether it pops up with 'wild guess' a lot; wouldn't be surprised! Ditto the 'I likes'. Question – if they just put the past tense in without thinking (consistent with what Flora said about English writing) then how can we expect them to translate accurately? They're not metacognitive enough with tenses...

15.7. Examples of paused translation mark scheme

Test AS7:

		Accept	reject	
	J'adore le skate	I love skateboarding	I like / anything that isn't skateboarding (eg roller / ice	3
2.	et j'aime aussi le rugby	And I also love rugby		5
3.	Moi, j'aime le tennis	(me I like rugby). Accept also without 'me'. Love also OK		3
4.	et j'adore la danse	And I love dance / dancing	I like	4
5.	J'adore le vélo	I love cycling / bike-riding	My bike	3
6.	et aussi le foot	And also football	foot	3
7.	Moi, j'aime la gymnastique	Me, I like gymnastics (omission of 'me' is acceptable)		3
8.	et j'adore la danse.	And I love dancing	Ignore as was given before	3
9.	J'aime le judo et les jeux vidéo	I like judo and video games	Anything else to do with videos (eg. watching videos)	5
10.	Moi, j'adore le vélo et la musique	Me I love cycling and music		5
11.	J'adore le rugby et les jeux vidéo	I love rugby and video games	Ignore the video games as it was given before	3
	TOTAL			34

Test F25 illustrates some of the nuances of the 'one mark per word' scheme – which language? Marking included qualitative interrater ('NRR' is my interrater) where we grappled with these ideas to develop logic and consistency.

		Accept	Details	points
1	Je joue du piano			3
2	Moi j'aime jouer de la guitare		Me, I = 1 Like to <u>play / playing</u> = 2 Guitar = 1	4
3	Je ne joue pas d'un instrument		I don't play = 3 An instrument = 2	5
4	Mais je suis chanteur dans un groupe	<u>Question about article point in je suis chanteur – NRR is happy !</u>	But = 1 I am = 2 a/ the singer = 1 In a group / band = 3	7
7	J'adore écouter de la musique classique		I love = 2 To listen to = 1 <u>Classical music</u> = 2	5
8	Je préfère télécharger des chansons	<u>No point for article 'des' because we wouldn't in English</u>	I prefer = 2 <u>Downloading</u> = 1 <u>Songs</u> = 1	4
				28

More reflections on the nuances of marking: this time test E3

mes frères et soeurs sont formidables		
my brothers and sisters are great	6	0
my brother and sister are great (or similar)	5	1
my brother and sisters are great	5	1
my brother and sister are (mistranslation of form	4	2
my brother and sister(s) are	4	1
my brother and sister.	3	1
my brother and sisters are.	5	0
my brother.	1	1
my brothers and sisters.	4	0
my dad / father	0	2
my friend is	0	3
my sister / brother is formidable	2	2
I have a brother and a sister	3	4
<i>Other notes</i>		

GUIDING PRINCIPLE: if in doubt, marks are about what they heard in the French rather than their English rendition, because I'm interested in their processing of the French. Therefore mes frères et soeurs - they will get marks for brother and sister because it would sound the same. The thing they got wrong is 'mes' even though that translates as 'my' because they didn't pick up on the fact that actually it's my-plural

15.8. Rasch model underpinning statistics

The scores that began as percentages after marking the paused translations were converted into a scale of 0 to 10 and carried into an excel spreadsheet where each score was identified by a concatenated code consisting of the participant number, participate code, and time point.

This analysis puts lots of separate ‘exam’ items onto a single underlying trait of LC ability and we’re tracking people over time to see whether they get any better or not, and also to see whether the individual exam items 0 how easy or difficult they are relative to each other.

Data was then imported into RUMM, a specialist statistical software package designed to create Rasch models. The sample size and missing data was found to be workable, and a compromise was made between the two: any participant who had completed fewer than fourteen tests was removed altogether from the study at this point, leaving a sample size of 80 participants but a more data-dense data set (ie less missing data as a proportion). The aim was to be able to draw a progress graph for each of the 80 participants with a score which represented their progress in paused translation.

Starting with the 0 to 10 data, the RUMM output created a model with an excellent fit, giving me ‘logit scores’. In Rasch terminology, the ‘logit’ is the unit of measurement in the scale used. (I also tried a 0 to 5 scale but given the success of the 0 to 10 scale, I felt that 0 to 5 risked losing nuance in the data. I also tried a 0 to 100 scale, but this gave too much detail for the Rasch model to cope with.)

The successful iteration of the output modelled all 1452 entries (ie 80 participants x 20 time points, with 148 missing data points) on a logit scale that ran from -8.369 to 6.89.

The scale was then converted so that the lowest score was 0, as it does make sense to consider negative proficiency, by adding 8.369 to each logit score, to give a range that ran from 0 to 15.268. For future rolling correlations, these scores were finally converted into a scale from 0 to 100.

Table 55: Rasch analysis: Screenshots of data set-up to export into RUMM

RUMMID	participant code	Tim	A57	A54	A58	M2	F25	G78	E11	F9	F26	F28	E3	G32	E56	F46	S12	D8	E65	D25	D3	D4	C48	C17	E56	C55	D31	C2	C41	C11	S13	S132a			
71SHOE5	71 SHOE	5					5					6	2																						
72STERS	72 STER	5					2					0	1																						
73TAYJ5	73 TAYJ	5					3					5	1																						
74VECK5	74 VECK	5					3					4	1																						
75WALH5	75 WALH	5					9					7	4																						
76WALB5	76 WALB	5					5					9	5																						
77WAHA5	77 WAHA	5					6					8	4																						
78WARH5	78 WARH	5					4					3	1																						
79WEBD5	79 WEBD	5					5					9	5																						
80ZAHCS	80 ZAHC	5					4					3	1																						
1ANDJ6	1 ANDJ	6					4						8	7																					
2ANDC6	2 ANDC	6					8						9	7																					
3ASHS6	3 ASHS	6					5						5	5																					
4AZOM6	4 AZOM	6					7						6	7																					
5BARN6	5 BARN	6					4						6	6																					
6BARG6	6 BARG	6					4						1	4																					
7BARE6	7 BARE	6					7						8	5																					
67SAGS14	67 SAGS	14										1													1										
68SALA14	68 SALA	14										0												0											
69SARL14	69 SARL	14										6												5											
70SEWH14	70 SEWH	14										1												0											
71SHOE14	71 SHOE	14										5												3											
72STER14	72 STER	14										1												0											
73TAYJ14	73 TAYJ	14										1												0											
74VECK14	74 VECK	14										2												1											
75WALH14	75 WALH	14										4												2											
76WALB14	76 WALB	14										7												6											
77WAHA14	77 WAHA	14										8												5											
78WARH14	78 WARH	14										2												1											
79WEBD14	79 WEBD	14										4												2											
80ZAHCS	80 ZAHC	14										1												0											
1ANDJ15	1 ANDJ	15										8												5											
2ANDC15	2 ANDC	15										9												8											
3ASHS15	3 ASHS	15										5												5											
4AZOM15	4 AZOM	15										9												4											
5BARN15	5 BARN	15										6												1											
6BARG15	6 BARG	15										8												3											
7BARE15	7 BARE	15										0												2											
8BIRO15	8 BIRO	15										5												2											

Output viewed in Excel graphs can be seen in Figure 124, initially in logits, then on a 0 to 100 scale, for participant ‘ANDC’.

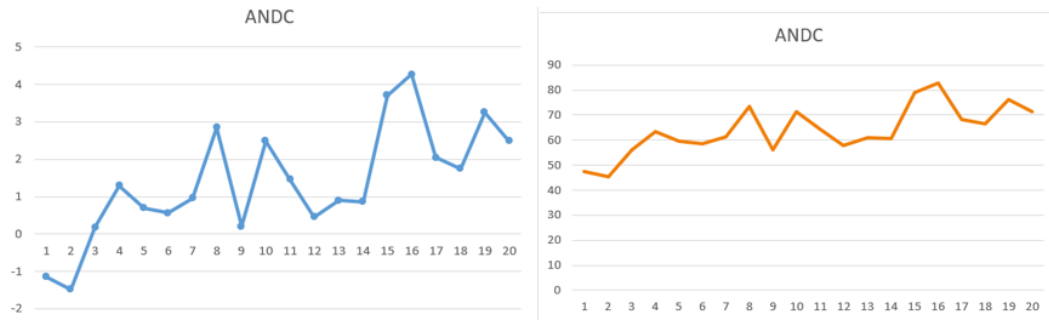


Figure 124: Rasch model outputs examples

15.9. Semi-structured interview protocol – sample from T11

1. How's things been? How's school? (Cs and Rs?)
2. How are you feeling about French now?
3. How's your new group (probe friendships and importance of them) and new teacher (mutual liking - is this important?) now you've been together for a term?
4. Tell me about listening in French at the moment.
5. If you had to rate the four skills in order of difficulty...
6. What about in order of preference?
7. When you're listening to a bit of French, can you talk me through exactly what's happening in your mind? (try to find out whether others are 'seeing words'... writing down what they hear in their mind's eye - and if so what lang are they writing in?) To what extent are they translating? When they say they're translating, are they actually doing something else?
8. How much are you understanding automatically? What about the bits you don't understand automatically - what's happening?
9. Tell me about what strategies you might use when you're listening to make it easier. Have you been taught those strategies or have you made them up?
11. Stimulated recall of D3 - Transcript
12. Read aloud.

Normalement, nous n'allons pas en vacances. Nous restons ici. Je déteste ça parce que c'est ennuyeux.

13. Elicited imitation (repeat T4). I'm going to say it and you're going to say it back to me. *Le matin, je fais du vélo et quelquefois je vais à la piscine avec mon frère.*

14. Can you tell me what you think it meant?

15. Is there anything else you'd like to add?

[my notes specific to Fran for T10, to pick up on] - liking to be pushed. Friends still very important in a class. Not done a lot of listening. Feels like all or nothing - either really hard, or really easy. And there are more that are really hard.

15.9.1. Typical interview transcript

Helen 17th July Time 6

Helen, you've done a year at Matthew Arnold!

[Big smile.]

Tell me about it. How's it been?

Good?

Yeah? What's been good?

Like lots of things and more confident at lots of subjects.

Yeah, Are you? When did that happen?

I'm not really sure.

Just slowly?

Yeah.

Because you did the play didn't you?

Yeah.

And you were the king?

Yeah.

Yeah. Does that help with your confidence?

Yeah, And at the end of year six, we did a play. And to practice my lines, I had to go into the playground and shout them at one of the teachers because they said you're not loud enough.

And Was that good?

Yeah.

You were louder then, were you?

Yeah.

Super. Okay. So you feel more confident.

Yeah.

And what about French? Just talk about your year of French.

At the start, like, I found it like, like, start I found that like, quite hard, but then it got easier. And then we like moved up levels, and then it got harder for like, because in Italian, we just learned like the really, really basic, like feminine and masculine stuff. We didn't really learn any of the tenses or anything. But like, it's got better cos I practice more. And I have this new app that my mum uses to learn German. called Duolingo. And my dad is learning French as well.

And are you learning French on it?

Yeah. And my brother.

Does it make you do really silly sentences, like, the cows do not drink the milk, things like that?

One of them that I had was I am a cat.

Okay. And are you?

No!

Okay. And so what about the French lessons?

Um, I think that they're good. And it took me a while to get used to like, people like that how they're like voices are when they speak in French for now. But now I've got it.

You've got a smile when you're talking about it. How do you feel about listening in French at the moment?

And I think that it's good. But in the test, at the start, I accidentally wrote in English instead of in French,

you told me that. Yes.

Because, like, and so I could hear like, like, the person next to me was asking a question, and I could hear like loads of English, so then I started typing in English.

Okay, was that just for one of them?

No, it was for that part? But I knew what the French was for it.

Okay. So should I just pretend you wrote in French for that and mark it like that?

Okay. And how do you feel about listening in French more generally? So not just the test that you did this morning, but

I think it's better and I'm really happy with my mark in the assessment.

What did you get?

Yes, I got a 5U.

Well done.

That's as well as you could have done.

Yeah. Excellent. How do you think you managed that?

Like, I told myself not to panic.

Did that work?

Yeah.

I think for some reason, it's harder with headphones on. Because with headphones. I feel like I'm going to miss something, or something. And then I, like, start panicking for some reason.

Does it feel like a little bit more serious, maybe?

Yeah.

Yeah. Okay.

But in the assessment, it's really quiet. And then I'm like, focusing? And like, some reason I feel like more like, like, I can hear it again, but with the French listening on the computer, you can only hear it once. So then I'm like, what if I miss this, and then what will I do?

So do you sort of overthink a little bit?

Yeah.

So you didn't panic in your listening assessment in class?

No, I was just like...

Not at all?

Not really

Okay.

Only on one of the sentences, I was like, wait, what? But then I got it.

Okay, super. And you must have got it all completely right.

Yeah.

Yay, hurry for you. So when you are listening in French what sort of problems are you having?

like, sometimes I'm busy translating one word and then I'm missing... Or, like, if I hear a word that I can't work out because they said it quickly or if I haven't heard it spoken

before, then I I'm like 'what'? and then I miss some of the other part. And then I don't do it as well as I could.

So. Say that again. You translate... if you're busy translating one word, you miss the next bit...?

Yeah. Like if I'm translating and then, or if I can't work out what a word is, then I focus on that too much and then I miss the rest of it.

And is that your main issue, do you think?

Yeah, - like being spooked out if I hear a new word or something? Then it throws me off a bit.

But you actually do hear the words, you'd be able to say you reckon that you can hear the gaps between the word?

Yeah.

Super. And then do you have any strategies to make it easier?

Well, I tell myself not to panic.

And does that work?

Yeah. Normally. Unless it's getting really hard. And then, like, if I couldn't do, one of the easier parts, then how will I do in the hard part?

So you're overthinking again?

Yeah.

Okay, so you have a little stern word with yourself, do you?

Yeah.

And that works, does it?

Normally

Any other strategies to make it easier?

Well, kind of, but sometimes I don't really notice that I'm doing them.

OK. Yeah. Fair enough.

The thing with memory is that I can normally remember something that's unusual but not something that's usual. Like if it's unusual it stands out, but if it's just a number or something...

Can you remember unusual words?

Yeah.

So if I told you that the German for a helicopter was a Hubschrauber – are you going to remember that for the next time I see you?

I'll try

we'll see. That'd be interesting.

Sometimes I'm too busy remembering one part and then I can't remember the rest!

Exactly! And that's exactly what we're trying to understand is how, how good you are doing that. And it's one of those things that you can't ever even get better at. It's just, it's like saying, you know, have you got blue eyes and brown eyes or something like that, or straight hair or curly hair...

Okay. So we're going to go back over one of the ones that we listened to. So it's this one here. This was the last one of the three that we did. So I'm just going to play it.

Tell me roughly what you think you got and how you got to the answer.

J'habite dans une station de ski à la montagne

What did you write and how did you get there?

I got a bit confused when I heard the 'station' or something. But when I saw it in the script, I knew that it was station. But originally I put that it was skiing in the mountains, that kind of thing.

So say what you wrote?

I can't remember, but something about skiing in the mountains.

Skiing in the mountains. And then you said – what did you say after that? I live there?

Yeah, but I can't remember exactly.

Because this was – do you remember – the introduction said it was people talking about where they live? So you think you wrote skiing in the mountains?

Yeah.

And how did the 'station' confuse you?

Because I thought it was statue when I heard it really quickly, so I was like, what? But then when I actually saw it in the script, I was like oh! Station.

So 'statue' threw you because why would there be a statue of skiing in the mountains, sort of thing?

Yeah.

Next one.

J'habite dans une petite ville

I live in a small village – I think I put – but it could be town.

What was the French word you heard?

Er, ville.

So the difficulty is remembering whether he was town or village?

Yeah, cos like, I think I put village. But then afterwards I was like, Wait, is it town? But I had already clicked it off.

Okay. And what do you think it is now?

I think it's town, but I'm not sure.

Okay, so that one was easy. Made sense?

Yeah.

didn't need any translating in your head?

No,

Just made sense in French. Next?

Dans l'est de la France avec ma famille

I think I put I live in France with my family, but I didn't understand the bit before it said France.

Dans l'est de la France ?

Yeah. I think... in something...

In? Go on? L'est de la France?

Like in the France, but I can't remember the l'est thing

So 'with my family', but that bit, you couldn't even segment it, it was just... was it a great big block?

Yeah.

L'est de la

Because even with the script thing where it shows you the words, I still didn't really know what that meant. Because I know it was, like 'est' is like 'is' and it's spelt the same, but I didn't know why there was an 'l' there.

Okay, next one.

Moi, j'habite avec ma mère

Um, I put, like.. I live – I can't remember – I live with my mum.

OK, and how did you get that?

Because, like, I heard him say 'ma mère' and I can't remember the rest of the sentence

But you're absolutely right, that's exactly what it said. I live with my mum.

Notre ville est grande et industrielle, et je ne l'aime pas.

For that one I only got, like, a bit of it but I can't remember which part. But when I saw it in the script again, I still wasn't sure what some of it meant – like 'industrielle' or something. Then I guessed and I thought it was industrial.

Yes, absolutely. Did you guess that when you saw it written down?

Yeah.

Not when you heard it.

No, because, like...

What did you write?

I can't remember.

Did you write anything at all, do you think?

Yeah, I definitely wrote at least two words. I can't remember.

And when you hear that now – it's probably difficult because you've seen the script, isn't it?

I can't remember – what did the script say? I can only remember, like, two words

No worries, that one was hard.

J'habite dans un port de pêche dans le sud de la France.

Um, I live... i put I live in a fishing village, but I don't know if that's right or not.

Where did you get that?

Cos pêche, I think, fish, and then it said j'habite which is I live.

And what about the end bit?

I can't remember what the end bit was.

Do you want it again?

Yes please.

J'habite dans un port de pêche dans le sud de la France.

I put in France again because I wasn't sure what that was. But I knew that it wasn't 'in' because it was like 'the'. So I put something about France. But I'm not sure.

OK. That sounds good.

J'habite dans une petite ville au bord de la mer

I think I put I live in a little village at the seaside? Because 'bord de la mer' is the seaside.

And, like une petite ville is like – I think I actually put small town because I realised from the toehr one that it was town, so I put town.

So spot on, and again it sounds like you didn't even need to translate that in your head; you were sort of understanding the French?

Yeah.

And then it gives you a little bit more information

Avec mon frère et mes parents.

With my brother and my parents.

Again, got it without translating it? Spot on.

I'm pretty strong on those two words.

That's brilliant. You know this is a whole GCSE question, as well?

No!

It is! Go you! That's pretty impressive, isn't it? You big clever clogs, you.

So you can really do it and can't you? It's just sometimes that the speed is too fast for your brain. But if we slow it all down, is dead easy.

And usually when I'm like speaking – especially in French – I can see the words in my head. But if I don't understand the words, then...

Then you can't see them in your head.

Yeah.

So you're sort of writing in your head quite a lot of the time?

Yeah.

And these days, you sort of writing it in French in your head?

Yeah, mainly. But in some sentences I write bits in English but I'm not so confident on, and then bits in French that I just know.

Cool. So that's moved hasn't it? from before? You were translating everything?

Yeah.

Okay. Interesting. Now, do I have any special questions for you?

Well I still feel like it's kind of a natural instinct to translate it. But I don't really see it if I'm translating – my head just goes [clicks fingers] like that and I understand the French.

And so does that make you feel a bit calmer about everything?

Yeah

Super. Now, last time, we talked about the possibility of watching some children's TV in French. Did you ever do that?

Oh yeah, I watched two episodes of Peppa Pig

How was it?

It was good. Yeah, one of them. I'd seen it before. And I was little and I remembered it. So I knew what most of it meant.

Well, that's really good, though.

Yeah,

Or did you feel that was cheating?

I felt like that was cheating a bit. But like, on the other one, like, never seen it before?

And how was it?

It was, it was good. And I could get some words from how they said it and stuff, even though so the elephants had like really strange accents.

Um, how's your speaking now? You still worried about speaking?

Well, I feel like for some reason, when I whisper it sounds better than when I say it out loud.

Okay.

Like, sometimes she's like, this is how you say it. And then she puts it on the board. And we have to look and then we'll do the repeating. I repeat it anyway but under my breath. And then it sounds really good. And then I say out loud and my voice sounds all shaky.

But are you feeling calmer in general?

Yeah.

I'm getting an impression that you're feeling much more confident.

Yeah.

Super. So is there anything else you want to add to me?

Not really.

You're going to France in the summer?

Yeah.

Are you all ready for your holiday?

Yeah.

When are you going?

We're going in a couple of weeks, to Marseille. Seven hours on the Eurostar. And then we've got a hotel with the Eurostar which is a really good price, and it has a pool.

In Marseille?

Yeah.

Oh lovely.

And the beach is close, so I'm really excited. Cos I've only ever been to about three or four beaches, in Wales. And never anywhere else, really.

And the sea will be lovely and warm, won't it?

Yeah.

Excellent.

That's why we went, because I was like, mum, can we go somewhere where the sea's warm, and she was like, OK, I'll think about it.

Excellent. How exciting. Will you tell me all about it when I see you in September?

Yeah. Definitely.

I'll let you get back to your English lesson.

15.10. Codebook summary

Name	Description	Files	References
Funny examples		10	12
interesting examples		35	55
double up two wrongs		4	4
false friend doubled up with real friend	Whereby a word is identified and translated twice, both accurately and with its false friend (e.g. <i>l'histoire c'est ma matière préférée</i> = history and maths are my favourite subjects)	19	27
some kind of error		60	22590
Grammar		58	2346
didn't detect plurals	Where a plural noun is given in the French but the participant has translated it in the singular. For example <i>'avec mes amis'</i> is translated as 'with my friend'.	25	625
hasn't processed verb ending	Where the English translation has detected the root verb but the meaning of the verb ending has been overlooked. This could be a tense issue (e.g. <i>j'ai joué</i> is translated as 'I play') or a modals issue (<i>Je veux jouer</i> is translated as 'I play') It will largely overlap with other categories, mostly 'wrong tense'.	37	858
incorrect negative added	Where a French sentence is given in the affirmative but the participant has translated it with a negative (e.g. <i>je joue</i> = I do not play).	24	73
incorrect plural given	Where the French gives the noun, plus any related articles, in the singular, but the participant translates it as a plural (e.g. <i>avec mon ami</i> is translated as 'with my friends').	14	74
missed negative	Where the French includes a negative formation of the verb (ne pas / ne jamais etc) but the participant correctly translates the verb and omits the negative (e.g. <i>je ne joue pas</i> is translated as 'I play').	14	86
wrong tense	Where the French audio is in a specific tense but the participant's translation does not reflect this (eg <i>je vais en France</i> is translated as 'I went to France'). In many cases this might overlap with 'hasn't processed ending'.	32	630
Segmentation		59	2722

Name	Description	Files	References
got end, missed start	Where the participant has given at least the final two words at the end of the track correctly, but has not translated any of the beginning element. For example <i>'je vais au parc avec mes amis tous les weekends'</i> = 'every weekend with my friends'. At least two ideas need to be missing at the start. If there is only one idea missing, it would count as 'word omitted' under vocabulary. If only the single initial word / idea is given, this is 'single idea recognised' under vocabulary'.	52	388
got start, missed end	Where the participant has given at least first two words at the end of the track correctly, but has not translated any further. For example <i>'je vais au parc avec mes amis'</i> = 'I go to the park'. At least two words need to be missing at the end. If there is only one word missing at the end, it would count as 'word omitted' under vocabulary. If only the single initial word / idea is given, this is 'single idea recognised' under vocabulary'.	59	1669
heard start and end; missed middle	As the categories above, but here there is a gap in the middle of the text. It needs to be more than one word / idea, otherwise it would be 'word missing' under vocabulary.	32	147
overwhelm		0	0
word boundaries not found lead to incorrect meaning	Where incorrect translations have arisen due to a failure to detect word boundaries. This might be a false friend (e.g. 'ski nautique' = 'skin') or the understanding of a different French word (e.g. 'plein d'espace pour' = 'passport'; <i>'ma petite sœur'</i> = 'my pizza).	48	502
liaison		6	21
word boundaries same letters	e.g. <i>musique classique</i> - repetition of /c/ sound	3	11
Single ideas only		59	2725
recognised cognate only	Where a single word or idea was translated from the track, and this was a cognate. For example <i>'je joue au football dans le parc avec mes amis'</i> = 'park'	55	972
recognised single idea only	Where a single idea only was identified in the track. For example <i>'Nathan, le plus petit de la famille, est très amusant'</i> = 'funny'.	57	1753
Vocab		60	8691
word missing	Then track is correctly translated except that the participant has left out a single word. For example <i>'j'écoute de la musique rap'</i> = 'I listen to music'.	60	3412

Name	Description	Files	References
content word missing	Then track is correctly translated except that the participant has left out a content word. For example ' <i>J'écoute de la musique rap</i> ' = 'I listen to music'.	60	1735
function word missing	Then track is correctly translated except that the participant has left out a single word. This could be initial ' <i>Mais il est sympa</i> ' = 'he is nice' or midstream ' <i>tu es vraiment difficile</i> ' = 'you are difficult'	46	1666
initial function word missing		27	1054
midstream function word missing		36	597
word wrong	Here, the sentence is translated correctly with the exception of one incorrect word. For example, ' <i>Je préfère la natation</i> ' = 'I prefer' tennis. sub-categories below.	60	5279
content word wrong		60	4991
'I like' for a different verb	Note: not a single word offering. Here there must be a full sentence but with the verb 'to like' or 'to love' in place of the verb given. E.g. ' <i>je joue au foot</i> ' = 'I like football'.	52	748
textual false friend		42	495
favourite-prefer		22	272
Phonological false friend		58	1057
wrong family-friends word		22	286
frère friend		16	76
wrong pronoun	Where pronoun was mistranslated (e.g. 'il joue = 'I play'). In some cases, this might also be coded under grammar to 'didn't process verb ending (e.g. ' <i>nous jouons</i> ' = 'I play'.	49	706
wrongly recognised single word only	Where a participant has given a single word translation of the track they have heard, but the word does not figure in a correct translation. For example ' <i>Elle n'habite pas loin de Bordeaux</i> ' = 'and'	50	337
function word wrong		42	225
wrong time phrase		30	182
wrong language		7	13

15.11. Screenshots demonstrating the evolution of coding

15.11.1. Codebook 3.1.23

Name	Files	References	Created On	Created By	Modified On	Modified By
Vocab	0	0	09/12/2022 12:01	KS	14/12/2022 17:09	KS
word missing		1	14/12/2022 17:10	KS	14/12/2022 17:54	KS
function word missing		0	30/11/2022 17:28	KS	14/12/2022 17:11	KS
midstream function		2	56 02/01/2023 17:35	KS	03/01/2023 11:51	KS
initial function word		2	199 02/01/2023 17:34	KS	03/01/2023 14:32	KS
content word missing		3	141 13/12/2022 12:32	KS	03/01/2023 15:15	KS
recognised cognate only (po		2	28 30/11/2022 17:47	KS	03/01/2023 15:07	KS
word wrong		2	40 09/12/2022 12:19	KS	14/12/2022 17:10	KS
function word wrong		2	9 14/12/2022 17:11	KS	03/01/2023 15:02	KS
aussi-assez		2	13 02/01/2023 17:54	KS	03/01/2023 12:06	KS
content word wrong		3	188 14/12/2022 17:11	KS	03/01/2023 15:11	KS
Grammar		0	0 09/12/2022 12:01	KS	14/12/2022 17:09	KS
didn't detect plurals		1	1 14/12/2022 17:13	KS	02/01/2023 17:01	KS
incorrect negative added		2	27 02/01/2023 17:33	KS	03/01/2023 15:03	KS
Strategy		0	0 09/12/2022 12:02	KS	14/12/2022 17:09	KS
Segmentation		1	11 09/12/2022 12:01	KS	14/12/2022 17:49	KS
heard start and end; missed		2	27 30/11/2022 17:33	KS	03/01/2023 15:00	KS
got end, missed start		3	45 13/12/2022 12:38	KS	03/01/2023 14:36	KS
got start, missed end		3	80 13/12/2022 12:36	KS	03/01/2023 15:14	KS

15.11.2. Codebook 2.4.23

Name	Files	References	Created On	Created By	Modified On	Modified By
Funny examples		8	10 07/01/2023 19:27	KS	17/03/2023 12:57	KS
interesting examples		18	22 14/01/2023 16:42	KS	20/03/2023 12:57	KS
double up two wrongs		3	3 31/01/2023 19:14	KS	04/03/2023 18:51	KS
false friend doubled up with real friend		14	19 17/01/2023 17:45	KS	20/03/2023 16:22	KS
Grammar		40	863 09/12/2022 12:01	KS	31/03/2023 20:33	KS
missed negative		4	6 01/02/2023 19:02	KS	22/03/2023 17:57	KS
incorrect plural given		9	47 13/01/2023 20:07	KS	22/03/2023 17:57	KS
incorrect negative added		16	60 02/01/2023 17:33	KS	22/03/2023 17:57	KS
wrong tense		23	166 01/02/2023 20:09	KS	22/03/2023 17:57	KS
didn't detect plurals		23	584 14/12/2022 17:13	KS	22/03/2023 17:57	KS
Segmentation		53	1736 09/12/2022 12:01	KS	22/03/2023 18:29	KS
Strategy		88	3512 09/12/2022 12:02	KS	22/03/2023 17:23	KS
CASE STUDIES		49	294 14/01/2023 18:23	KS	22/03/2023 17:23	KS
guess		42	632 01/02/2023 11:33	KS	31/03/2023 20:38	KS
Context		58	2586 03/01/2023 18:29	KS	22/03/2023 17:23	KS
Vocab		56	7374 09/12/2022 12:01	KS	31/03/2023 20:24	KS
wrongly recognised single word only		30	95 14/01/2023 17:03	KS	22/03/2023 18:29	KS
recognised false friend only		30	150 13/01/2023 19:46	KS	22/03/2023 18:29	KS
recognised cognate only (poss strat)		39	409 30/11/2022 17:47	KS	22/03/2023 18:29	KS
recognised single word only		39	658 07/01/2023 19:48	KS	22/03/2023 18:29	KS
word missing		42	2812 14/12/2022 17:10	KS	22/03/2023 18:29	KS
word wrong		50	3249 09/12/2022 12:19	KS	22/03/2023 18:29	KS

15.11.3. Codebook 30.5.23

RQ1 first attempt.nvp - NVivo 12 Pro

Name	Files	References	Created On	Created By	Modified On	Modified By
phonology		0	17/05/2023 21:53	KS	17/05/2023 21:53	KS
voice onset		0	17/05/2023 21:53	KS	17/05/2023 21:53	KS
consonant salience		0	17/05/2023 21:53	KS	17/05/2023 21:53	KS
Funny examples		10	12/07/01/2023 19:27	KS	26/05/2023 20:26	KS
interesting examples		35	55/14/01/2023 16:42	KS	28/05/2023 14:17	KS
double up two wrongs		3	31/01/2023 19:14	KS	04/03/2023 18:51	KS
false friend doubled up with real friend		19	27/17/01/2023 17:45	KS	28/05/2023 11:41	KS
Grammar		64	2351/09/12/2022 12:01	KS	30/05/2023 18:32	KS
incorrect plural given		13	72/13/01/2023 20:07	KS	30/05/2023 18:32	KS
incorrect negative added		24	73/02/01/2023 17:33	KS	30/05/2023 18:32	KS
missed negative		14	86/01/02/2023 19:02	KS	30/05/2023 18:32	KS
didn't detect plurals		29	630/14/12/2022 17:13	KS	30/05/2023 18:32	KS
wrong tense		34	632/01/02/2023 20:09	KS	30/05/2023 18:32	KS
hasn't processed verb ending		37	858/27/04/2023 12:48	KS	30/05/2023 18:32	KS
Segmentation		74	2602/09/12/2022 12:01	KS	30/05/2023 18:32	KS
overwhelm		2	3/14/01/2023 18:39	KS	30/05/2023 18:32	KS
complaints about speed		10	11/14/01/2023 18:36	KS	30/05/2023 18:32	KS
heard start and end; missed middle		33	147/30/11/2022 17:33	KS	30/05/2023 18:32	KS
word boundaries not found lead to inco		48	364/14/01/2023 16:39	KS	30/05/2023 18:32	KS
got end, missed start		52	385/13/12/2022 12:38	KS	30/05/2023 18:32	KS
got start, missed end		61	1679/13/12/2022 12:36	KS	30/05/2023 18:32	KS

Name	Files	References	Created On	Created By	Modified On	Modified By
Vocab		77	11164/09/12/2022 12:01	KS	30/05/2023 18:32	KS
wrongly recognised single word only		51	337/14/01/2023 17:03	KS	30/05/2023 18:32	KS
recognised false friend only		51	449/13/01/2023 19:46	KS	30/05/2023 18:32	KS
recognised cognate only (poss strat)		60	981/30/11/2022 17:47	KS	30/05/2023 18:32	KS
recognised single word only		60	1755/07/01/2023 19:48	KS	30/05/2023 18:32	KS
word missing		63	3416/14/12/2022 17:10	KS	30/05/2023 18:32	KS
function word missing		46	1666/30/11/2022 17:28	KS	30/05/2023 18:32	KS
content word missing		63	1739/13/12/2022 12:32	KS	30/05/2023 18:32	KS
word wrong		71	4225/09/12/2022 12:19	KS	30/05/2023 18:32	KS
function word wrong		42	223/14/12/2022 17:11	KS	30/05/2023 18:32	KS
content word wrong		69	3941/14/12/2022 17:11	KS	30/05/2023 18:32	KS

15.12. Ethics paperwork

CUREC

LG Liam Gearon <liam.gearon@education.ox.ac.uk>
Wed 18/07, 19:13
Kedi.simpson@education.ox.ac.uk; Jackie Bridges; Education Research Office; Jess Briggs

Inbox

Dear Kedi,

The developmental trajectory of second language listening

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I note the school is named in this application, please ensure strict anonymity.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Please note data should not be collected until approval granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse
<http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf>

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Sincerely,

With kind regards,

Liam

Dr LF Gearon
Chair DREC
Department of Education
University of Oxford

Re: CUREC 1A Application - Kedi Simpson (DPhil) - Google Chrome

<https://outlook.office.com/owa/projection.aspx>

Reply | Delete | Junk | ...

Re: CUREC 1A Application - Kedi Simpson (DPhil)

NW Nicola Warren-Lee <nicola.warren-lee@education.ox.ac.uk>
Wed 02/05, 11:55
Education Research Office; kedi.simpson@education.ox.ac.uk; Robert Woore; jessica.briggs@education.ox.ac.uk

Inbox

CUREC_1A Kedi Simpso...
195 KB

Show all 1 attachments (195 KB) | Download | Save to OneDrive - Nexus365

Dear Kedi

Title and reference number:
The developmental trajectory of second language listening: PILOT
Ref: ED-C1A-18-190

The above application has now been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse http://www.nspcc.org.uk/Inform/research/questions/reporting_child_abuse_wda74908.html

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Best wishes
Nicola

15.13. Criteria for exclusion from study

Time 1: All valid!

Time 2: No LC answers' 'Bob Brushington' (unidentified participant), No answers from WA, SM, LS, EM

Time 3 No LC answers from WA, SM, LS

Time 4: No LC answers WA, SM, LS

Time 5: Unfinished: AM, BK, SM, BM, EM

Time 6 No LC answers: WA, MA, FK, SM, LR, LS

Time 7: No LC answers: WA, MA, KC, BK, SM, AM. LR, LS, TT

Time 8: No LC answers: TT, SM (minimal), LS, BK, WA, AJB

Time 9: No LC answers: MA, BK, BM (minimal)

Time 10: At this point participants began claiming to give answers, but without listening.

Total audio running time 860 seconds. All participants completed it in more time than that, although SM only 868 seconds. No LC answers: WA, BM, AM (minimal answers), LS (wrote 'basketball' for everything!)

Time 11: Total audio running time 816 seconds. LB completed survey in 376 seconds. No LC answers: WA, MA, GB – unfinished and most of what he did put was question marks! FG, BK, AS

Time 12: Audio running time 652 seconds. everyone did it in more than 652 seconds! No LC data: WA, LB, FG, BM, LR (only two answers), AS, LS.

Time 13: Audio running time 790 seconds. MA, FG completed in less than 700 seconds.
LC no data MG. BM, LR, AS, LS.

Time 14: Audio running time 772 seconds. The following 'completed' in under this time:
MA, LC, SdL, FG, RG, BK, BM, SM, LR, LS. No response in LC: GB, AM, AS

Time 15: Audio running time 702 seconds. The following 'completed' in under this time:
MA, SdL, SS, FG, EH, BK, AP, LR, LS. No response in LC: LB, SM, AM, AS JW, MG.

Time 16: audio running time 852 seconds. The following 'completed' in under time: MA,
GB, JC, HE, EH, BK, EM, EM, SM, AP, JW. No LC response from SL, MG, LR, LS

Time 17: total audio running time 851 seconds. The following 'completed' in under this
time: MA, LB, PB, HE, EH, BK, SM, ES, FK, MH, DL. No LC data GC, MG, EM, AM,
LS, MW

Time 18: Total running time 860 seconds. 5 students excluded for being faster than this.
additional for giving no LC data MA, FK, LS, ES, RC, LR

Time 19: total running time 791 seconds. Removed for completing faster than that MA,
LB, KC, EH, JH, SM, LS, RS. No LC data BK, BM.

Time 20: Total running time 898. Faster than this = LB, PBP, HE, EH, BM, EM, DR, LS.
No LC data SdL, EH, FK, EM, AM

15.14. Summary of test difficulty levels

I assumed that participants would attempt what they could, and therefore, the fewer attempts were made at translating a track, the harder that track was. To calculate overall test difficulty, therefore, I counted the maximum number of entries that could be made in that test by multiplying the number of participants at that time point by the number of

tracks in the test (e.g. if there are 10 tracks and 70 participants, there is a maximum of possible 700 entries.) For each test I then counted the number of actual attempts at each track, and divided them by the number of possible attempts and multiplied by 100 to give a percentage. Details are in Table 56. Counting was done manually. Figure 125 gives a taste of the kinds of entries that are made that would count as ‘not attempted’; it was not possible to programme NVivo to recognise what was effectively a non-attempt. My scores are as follows:

Table 56: Engagement scores of all listening tests

engagement data					test 1	test 2	test3	AVERAGE
	test 1	test 2	test3					
T1	as7	as4	as8	T1	89.7	86.9	79.5	85.4
T2	as8	stM23B	F25	T2	83.9	66.2	68.0	72.7
T3	G78	stM23B	E111	T3	73.3	67.1	59.1	66.5
T4	F9	AS7	E26	T4	71.3	93.8	61.1	75.4
T5	F28	F25	E3	T5	82.3	72.5	64.0	73.0
T6	G32	ST32B	E56	T6	74.0	67.1	71.3	70.8
T7	F46	F9	E26	T7	66.1	70.0	71.4	69.2
T8	ST2M12A	E3	D8	T8	80.6	61.3	71.0	71.0
T9	F25	E65	D25	T9	84.3	65.8	70.4	73.5
T10	E111	F9	D3	T10	82.2	70.7	66.3	73.1
T11	E3	D47	C48	T11	78.8	76.3	60.4	71.9
T12	E26	D8	C17	T12	72.8	77.6	72.7	74.4
T13	E56	D25	C55	T13	82.5	63.7	74.6	73.6
T14	E111	D31	C48	T14	77.0	70.2	62.0	69.7
T15	E26	D3	C2	T15	75.3	69.4	69.2	71.3
T16	E3	D8	C41	T16	77.4	82.1	72.7	77.4
T17	D47	C48	C11	T17	78.3	61.9	75.2	71.8
T18	D25	E111	ST3AP	T18	67.6	89.3	76.7	77.9
T19	E56	D3	ST351A	T19	82.3	69.5	72.2	74.7
T20	C48	ST342A	D8	T20	69.6	74.0	85.3	76.3

ID	On va toujours à une auberge de jeunesse
18	I'm not sure, sorry!
19	skipped
20	i don't understand
21	i have no idea
22	...
23	Don't understand
24	?

Figure 125: examples of 'not attempted'