

A glimpse into the expert lexicon: Behavioural evidence from Aviation English

Hilary Wynne^{a,*}, Henry Emery^b

^a University of Oxford, UK

^b Latitude Aviation English Services, UK

ARTICLE INFO

Keywords:
Expertise
Language
Processing
Aviation

ABSTRACT

Using a specialised language is often an integral part of being an expert in a field (e.g. medicine, aviation). We employed an auditory sentence processing task with immediate recall to compare behavioural (RT, accuracy) responses by experts (pilots with more than 200 hours of training, $N = 42$) to non-experts (listeners with no training in the field, $N = 54$). Auditory stimuli were purpose-built to reproduce the standard aviation operating environment: recordings were produced by an aviation professional employing authentic radiotelephony cadence and mixed with simulated control-tower background noise.

As expected, reaction times showed a clear effect of expertise: experts responded significantly faster on matched than mismatched trials, whereas non-experts showed no reliable effect. More importantly, accuracy analyses revealed that experts demonstrate strong sensitivity to phraseology, with significantly lower accuracy on recall tasks after hearing a single nonstandard word in an utterance than in all other conditions. Taken together, our results indicate that experts in aviation radiotelephony are principally sensitive to non-standard phraseology. These findings have implications not only for the organisation of the expert lexicon, but also the importance of using correct phraseology in the standard aviation environment.

1. Introduction

Expert knowledge refers to the deep, specialised understanding and proficiency that an individual has in a particular field or domain. The effects of expertise on mental perception and processing have been studied in many domains, including computer coding (Adelson, 1984; Kuo & Prat, 2024), chess (de Groot, 1965; Simon & Chase, 1973; Simon & Gilmarin, 1973; Gobet, 1998; Gobet & Simon, 1998), physics (Chi et al., 1981), mental calculation (Staszewski, 1990), medical expertise (Norman et al., 1989; Rikers et al., 2002; Schmidt & Boshuizen, 1993), astronomy (Bryce & Blown, 2012), and music (Maturi & Sheridan, 2020). Findings have agreed that experts and non-experts differ in the organisation of their knowledge: this includes asymmetries in the amount of subject-specific knowledge, the structure of this knowledge, and the strategies and procedures they apply when using this knowledge (cf. Kalverkämper, 1983, 1996). In domains such as chess, this is seen as complex knowledge “chunks ... patterned clusters of pieces... stored in

long term memory” (Gobet & Simon, 1998). While a growing body of research in cognitive science and psychology (cf. Hoffman, 1998; Kozma, 2020; Murphy & Wright, 1984) has focused on the difference between experts and non-experts, comparatively little has been done on the contribution of language to expertise.

In addition to specialised knowledge for procedures and concepts, Kalverkämper (1996) lists specialist communication (“*Fachkommunikation*”) and skills in communicative habits (“*Kommunikationsgewohnheiten*”) as a major part of domain-specific knowledge.¹ Indeed, using and comprehending a specialised language is frequently an integral part of being an expert in a field (see, for example, aviation and marine operations). Often this language takes the form of an ESP (English for Specific Purposes) and is taught with a focus on the specific needs of learners within a particular field or occupation rather than general language skills: this regularly requires the speaker to activate, understand, and use special linguistic features such as smaller vocabularies, distinct lexical classes, grammatical restrictions, and formulae

* Corresponding author at: University of Oxford, Faculty of Linguistics, Philology and Phonetics, Language and Brain Laboratory, 37 Wellington Square, Oxford, OX1 2JF, UK.

E-mail address: hilary.wynne@ling-phil.ox.ac.uk (H. Wynne).

¹ ... (b) speziell (durch Ausbildung, Lehre, Studium) erworbene Kenntnisse; (c) Fertigkeiten (in Arbeitsprozessen, spezifischen Handlungsabläufen, Umgangs- und Kommunikationsgewohnheiten); (d) (Erfahrungs- und Lern-) Wissen (zu einem Sachgebiet bzw. Handlungszusammenhang) (1996: 2–3).

expressions (cf. Gotti, 2008). An important question remains: what does this expertise look like in terms of the language faculty?

Most research on language processing has examined how speakers and listeners respond to deviations from expected structure in general language (e.g. grammatical errors, semantic anomalies, or pragmatic violations). These studies show measurable effects on behavioural responses, including longer reaction times and reduced accuracy, indicating increased cognitive load or processing effort. However far less is known about how deviations in domain-specific languages are processed. Limited research has indicated that experts who learn a specialised language are specially cued to disruptions in those languages' structures. For example, Kuo and Prat (2024) found that, the more training that coders had in the programming language Python, the greater sensitivities they showed to violations in code. Specifically, violations to syntax in snippets of Python code triggered the P600 response, an event-related potential (ERP) consisting of positive deflection in the 500–800 ms post-stimulus window that reflects a violation of syntactic expectations. This effect was stronger in Python programmers with greater expertise, which the authors took to suggest that syntactic rules for expert languages become more deeply internalized with experience. What remains to be investigated, however, is the effects of specific types of deviations- namely, deviations that are acceptable in general language but invalid within a specific domain.

One of the most characteristic features of a specialised language is monoreferentiality: that is, while words (and phrases) can have a wide range of meanings in regular discourse, they are only allowed a single referent (i.e. meaning) in specialised languages (Gotti, 2008). An example of this characteristic can be found in the Standard Marine Navigational Vocabulary, where the word “foul” can only refer to the twisting of a cable, line, or rope around the anchor or propeller. e.g. “my anchor is foul”. In this specialised language, “foul” cannot be replaced by synonyms (e.g. “tangled”) or paraphrasing (e.g. “all wrapped up”) (cf. Strevens & Johnson, 1983). This creates an important but unexplored question: how do listeners cognitively respond to words or phrases that are grammatically or semantically valid in everyday language, but have no place existing in the strict syntactic and semantic structures of specialised languages? Deviations to standard language occur regularly within the expert domains; after all, a vocabulary of 300–400 words (such as that in aviation) can only cover so many possible situations. By examining how deviations influence behavioural measures during comprehension and task performance, we aim to be able to isolate the role of domain-specific expectations and learned linguistic constraints.

In this study, we employed a sentence processing paradigm combined with word recall in order to investigate the effects of violations of monoreferentiality in an expert language, Aviation English. The questions we ask are, (1) do experts and non-experts process violations in specialised languages differently as shown by explicit differences in our two behavioural measurements (reaction time and accuracy), and, if so, (2) what are the effects of violations of monoreferentiality on these measures in experts? Investigating the effects of deviations that are anomalous and/or incongruent *only within a specific domain* can reveal how domain expertise, contextual rules, and specialised conventions shape language processing- this has implications not only for theories of expertise, but also language comprehension.

2. Experts and non-experts: experimental evidence

The literature is fragmented across fields and, as a result, no unified picture for the specific architecture of the faculty for expert language exists. Recent studies have targeted the expert's vocabulary breadth, depth, or acquisition through paradigms designed to measure semantic memory and associative processes rather than lexical activation (cf. Ballot et al., 2024; Barbero & Amaro, 2024; Kim et al., 2011; Masrai et al., 2021; Vintar & Saksida, 2023). For example, Kim et al. (2011) found that, in written evaluation of events involving the individuals' domains of expertise, experts tended to use a wider range of nouns,

including general nouns, pronouns, and personal pronouns, while novices relied more heavily on adjectives and verbs. In addition, experts produced significantly more words per sentence. Vintar and Saksida (2023) conducted a word association task on experts and non-experts of karstology (the study of karst and cave landscapes) and found that experts were regularly more specialised in their descriptors (e.g. using the specialised terminology *karst* instead of the more generic *cave*), suggesting that they form specialised lexical entries that have highly compact domain-specific knowledge.

A limited number of psycholinguistic studies directly compare how experts and non-experts retrieve the same set of words. Wiedenbeck (1985) found that expert FORTRAN programmers exhibited faster response times and more accurate responses than non-experts in two different tasks: 1) a task requiring the detection of syntactic errors (e.g. *IF(X = Y) GO TO 1000 for IF (X .EQ. Y) GO TO 1000) and 2) a decision task involving descriptive phrases (e.g. Assignment statement), followed by a Fortran code segment (e.g. F = F + TOT), in which the participant judged whether the phrase was a correct description of the code. Postal (2004) found that proficient soccer players and trainers exhibited higher reading span scores than non-experts (novices and supporters) for ‘structured’ sentences (i.e. sentences containing high levels of contextual information) in which a critical word was related to soccer (e.g. for this match, the trainer has chosen Julien to play defence) compared to ‘unstructured’ sentences (e.g. what a beautiful reward for Sebastien's team to win the cup). Toth et al. (2019) tested three groups of Counter-Strike gamers (low, intermediate, and elite) and also compared them to a group of non-gamers using the Stroop Task. They found that elite gamers responded faster and more accurately than intermediate-ranked gamers, and all gamer groups (regardless of rank) were faster than non-gamers.

There is also a growing body of neurophysiological evidence for differences between experts and non-experts (or high- vs low-knowledge participants) in how domain-specific language is processed. Using EEG (electroencephalography), Bergum et al. (2024) used EEG combined with eye-tracking to examine brain activity in programmers with different levels of self-reported experience. Stimuli consisted of Java source-code snippets. Eye-tracking results indicated that proficient programmers exhibited shorter fixations on fewer code elements, suggesting that they processed the source code more efficiently. EEG data from time-frequency analyses indicated that programmers with higher efficacy exhibited lower beta and higher alpha and gamma power, leading the researchers to argue that cognitive load during the task for the expert group was lower than that of programmers with less experience.

Troyer and Kutas (2020) investigated N400 responses to violations in ‘Harry Potter knowledge’ (i.e. knowledge for the narrative world in the Harry Potter books) and found that the degree to which related violations elicited reduced N400 responses was modulated by an individual's degree of domain-specific knowledge. In addition to the P600, Kuo and Prat (2024) elicited N400 responses for semantically implausible code in their study, which they argue suggests that programmers generally rely on contextual plausibility as they visually process code: crucially, this occurred across all levels of expertise. Walla et al. (2024) found differences in event related potentials (ERPs) in legal experts and non-experts; critically, these differences were related to items that were considered deviations to a standard stimulus (i.e. law-relevant vs. law-irrelevant). Specifically, they found increased negativity for law-irrelevant lexica in experts ~450 ms post stimulus in both parietal and fronto-central (with left-hemisphere dominance) regions. This effect was absent in the non-experts.

Thus, while sparse, the psycholinguistic evidence has suggested that there are depths of lexical representation, and these depths shape how efficiently domain-specific words are accessed and how they compete with general-language alternatives. Existing neurolinguistic research has also underlined some differences in experts and non-experts, particularly in regard to expert knowledge (especially domain-specific referents and conceptual systems), which appears to leave measurable

neural traces both in ERP and time frequency analyses. Taken together, findings indicate that expertise involves not only the acquisition of new terminology with specialist semantic networking, but also the development of parallel, context-sensitive representational pathways for these items in the lexicon. However, there remains a lack of experimental evidence about precisely about how experts and non-experts disentangle lexical representation from general cognitive skills (attention, working memory, predictive processing) in expertise-related behavioural differences.

This study examines the interaction between specialised and general language systems in expert speakers and listeners, with a particular emphasis on how these systems are accessed and coordinated in real time. To address this gap, the present study employs an experimental environment designed explicitly to foreground linguistic processing and minimize the contribution of non-linguistic influences such as conceptual problem-solving, domain-specific inference, or physical proficiency. This approach enables a more direct examination of how experts navigate, activate, and integrate both lexical systems when processing stimuli that point to the same referents but differ in their degree of specialisation.

3. Expert language: the case of Aviation English

Expert languages have been given a wide variety of names: restricted languages, specialised discourse, sublanguages, domain-specific languages, Languages with Specific Purposes, prescribed languages and/or standardized languages (cf. Firth, 1957; Gotti, 2008; Harris, 1982; Kittredge, 1982; Wallace, 1981). Importantly, they are “microcosms of whole languages” (cf. Harris, 1982) that are categorized by the speakers/listeners who use them and the linguistic features they show. They have a dedicated community of speakers that share specialised knowledge and will often contain distinct word classes that reflect domain semantics, a restricted lexicon, special words, distinct lexical classes, restricted sentence syntax, domain-specific word classes, restricted text grammar, and a statistical (often reduced) profile of word occurrences (Kittredge, 1982).

The field of aviation places strict demands on clarity and safety and thus, a dedicated sublanguage has been established: Aviation English. It is a simplified, purpose-built subset of English designed to support efficient and unambiguous communication (cf. ICAO Doc 9835). Often referred to as “Radiotelephony”, or “Aviation English”, Standard Aviation English falls under the category of “English for Specific Purposes” (ESP). Aviation English is a highly specialised language used only by trained personnel on dedicated radio frequencies. These personnel draw on a shared body of professional knowledge gained through aviation training, including flight dynamics, navigation, weather terminology, mechanical systems, airport operations, and other procedural skills.

Because Aviation English relies so heavily upon the repetition of certain elements (altitude, heading, clearance instructions), the transfer of unproductive information is kept to a minimum. The sublanguage has a prescribed number and letter readout system and a very limited vocabulary (300–400 words). Aviation English also employs special words (*Roger*, *Charlie*), abbreviations (*ILS*, *VFR*), specialised jargon (*squawk*, *swing*), and elisions of longer phrases (*‘Wilco’* for *will comply*, *‘Ident’* for *identify*). Sentence structure in Aviation English is notably constricted, to the point where utterances would qualify as ungrammatical or deviant in natural language. See [Example 1](#) below, in which a controller is instructing an aircraft to enter the runway and hold position.

Example 1. *Shuttle Two Lima, Heathrow Tower, runway one five right, position and hold.*

Utterances in Aviation English not only exhibit random word order, but this order can change without the meaning of the utterance itself changing:

Example 2. [Cleared to land] [runway one niner right]
[Runway one niner right] [cleared to land]

This is due to the fact that each component (e.g. [cleared to land]) has a specific semantic meaning that can stand alone. In [Example 1](#) above, the controller’s instruction [position and hold] can only be associated with specific commands: the pilot of the departing aircraft is explicitly aware that they have only been instructed to manoeuvre **onto** the runway, not to begin take-off.

Monoreferentiality is critical in Aviation English. [Example 3a](#) below is a standard Aviation English utterance, in which a controller is instructing a pilot to climb to a specific altitude and remain there. It also infers that the pilot may not leave that altitude. Although the phrase ‘stay at’ ([Example 3b](#)) is synonymous with the word ‘maintain’ in Plain English, it can **never** be substituted in Aviation English.

Example 3. a: *Climb and MAINTAIN four thousand.*
b: *Climb and *STAY AT four thousand.*

Thus, word predictability in standard Aviation English is entirely preconditioned. MAINTAIN cannot be replaced with a synonym or paraphrase (as in 4 below):

Example 4. *Climb and MAINTAIN/ *STAY AT/ *GO TO/ *REMAIN.*

While this may seem excessively stringent, such an error can result in the failure of message transfer or worse. On December 16, 2012, a pilot was conducting an instrument approach to an airfield in South Carolina, USA when he experienced directional instrument failure. He advised the controller that he was “no gyro” (i.e. without gyroscopic flight instruments such as a compass) and requested an alternate airport which had better visibility and thus would be a safer option for an approach. However, the implications of “no gyro” were unclear to the air traffic controller, who directed the pilot to make another instrument approach—the pilot lost control of the aircraft on approach (NTSB ERA13FA088).

The structure of expert language (hierarchies, categories, abbreviations) reflects the way knowledge is organized cognitively. In aviation, an expert’s lexicon forms a tightly integrated network of terms linked to aircraft systems, operational procedures, and decision-making contexts. This system allows pilots to translate verbal input into precise, often automatic actions, reflecting deep domain knowledge and task fluency. Developing expertise requires learning specialised linguistic and textual structures that span multiple levels of language: these include (but are not limited to) specific sounds, words, formulaic phrases, and syntactic patterns.

At the same time, general linguistic knowledge acquired through everyday use remains active and can interfere, particularly when domain-specific forms conflict with more natural or familiar constructions. Evidence of this is seen when features of Plain English creep back into radiotelephony transmissions in a myriad of ways: minimizations, justifications, elaborate syntax, indirect speech acts, felicitous, lexical swaps, jargon, slang, misfires, anticipation and transposition errors all feature in daily exchanges in the air (cf. Wynne, 2025 for examples of non-standard language in aviation).

4. The study

When processing auditory language, we access entries in our mental lexicon and map their equivalent meanings onto the incoming speech signal (Schiller, 2021). The mental lexicon (cf. Aitchison, 2012) is a mental storehouse of words and their meanings that each person has learned throughout their lifetime: it contains information about a word’s phonology, meaning, orthography, morphology, grammatical properties, and how it relates to other words. As the brain receives phonemic information, it begins to search the mental lexicon for possible matches. This process is incremental and predictive: once a match is found in the mental lexicon, the brain accesses all relevant information about the word, and its meaning is integrated into the context of the utterance.

Evidence has shown that, when the signal contains an ambiguous word (i.e. a word that has more than one meaning, e.g. *bank*), the brain often activates all known meanings simultaneously or in rapid

succession (Bilenko et al., 2008; Onifer & Swinney, 1981; Swinney, 1979; Thompson-Schill et al., 1997). In selecting the correct interpretation, the context of the utterance (i.e. words, sentence structure, topic, tone, etc.) plays a vital role in determining which meaning is most relevant. Once the appropriate meaning is selected, the brain actively suppresses the irrelevant meanings to avoid confusion and maintain fluent comprehension; if context is weak or delayed, the brain may favour the more frequent or dominant meaning (e.g., *bank* as a place for money rather than a *bank* of a river). To investigate how this occurs in domain-specific language, we used a sentence processing paradigm combined with an immediate word recall task to probe how violations in specialised languages affect experts, and how their responses compare to non-experts. Specifically, our goal was to examine differences in behavioural responses (reaction times and accuracy rates) between experts (pilots with more than 200 hours of flight time) and non-experts (native speakers with no knowledge of aviation) when faced with standard and non-standard expert language (which was, in this case, Aviation English).

Thus, our first set of predictions centre around the expected differences between experts and non-experts. We predict that the responses will differ across groups due to the fact that Aviation English will obviously be unfamiliar to non-experts (and possibly even incomprehensible): this will undoubtedly affect the speed and accuracy at which participants respond. Our second set of predictions focussed on the expert group solely: in these speakers, we predicted that violations to standard Aviation English phraseology would make the word recall task harder to perform, resulting in asymmetrical reaction times and accuracy rates. The organisation of this study is as follows: first we present our experimental design and stimuli, followed by results for experts in Section 5 and results for the non-expert group (Section 6). We compare performances in a group-level analysis in Section 7, and discuss our findings in Section 8.

4.1. Design

The experimental stimuli consisted of 80 critical sentence pairs: 40 Standard Aviation English utterances (e.g. *Remain this frequency*) and 40 utterances containing non-standard phraseology (e.g. *Stay on this frequency*) semantically related to the expected exemplar. The immediate word recall task contained target words that were either matched or mismatched (e.g., *remain/stay*) to the point of violation in the utterance. There were also 80 filler sentences (e.g. *Descend and maintain 3000*), 40 of which contained unrelated semantic violations (e.g. *routine*). This resulted in 160 trials split across two lists. Our critical comparisons were between Utterance Type (Legal, i.e. Aviation English and Illegal, i.e. nonstandard phraseology) and Recall Condition (match/mismatch; see Table 1 for a full range of conditions).

Due to the nature of the standardized phraseology, lexical items in Aviation English (e.g. *remain*) had much lower frequencies than the Violation items (*stay*); they also mismatched for orthographic length (number of letters in the visual stimulus) (see Table 2). Conditions were counterbalanced across sessions such that participants saw an equal number of items in each session. Trials were pseudorandomized in order to distance similar utterances from one another and to avoid repetition of the same utterance type within one session.

Table 1
Example stimuli and conditions.

Stimulus	Utterance type	Recall word	Recall cond
Remain this frequency	Legal	REMAIN	match
		STAY	mismatch
Stay on this frequency	Illegal	STAY	match
		REMAIN	mismatch

Table 2
Recall word measures.

Utterance type	Recall condition	Freq (ppm)	Phono length	Ortho length	Cosine similarity
Legal	match	238.1	6.50	8.1	–
	mismatch	1981.4	6.51	5.0	0.29
Illegal	match	1981.4	6.51	5.0	–
	mismatch	238.1	6.50	8.1	0.29

4.2. Materials

All auditory stimuli were recorded by a monolingual British English-speaking Aviation English test designer in the cadence and structure of a typical ATC transmission, with computer-generated background noise added to simulate the standard operational environment. The intensity of all stimuli was also equalized to 75 dB, with an average signal-to-noise ratio of 22.2 dB. All trials followed the same design with each stimulus beginning with the same aircraft identifier, *Latitude 721*: this was done to simulate a transmission directed at the operating pilot.

4.3. Procedure

The experiment was administered on participants' computers using Pavlovia and coded in PsychoPy. Due to the large number of trials, the experiment was split into two parts of approximately 20 min each. Each session consisted of four blocks of 20 trials. In between each block, participants could take a short break. Participants first participated in a practice session consisting of 10 practice items in which feedback on correct responses was provided ('Correct/Incorrect'). For the experimental trials, there was no feedback provided.

Each trial (see Fig. 1) began with a fixation cross for 150 ms, followed by the auditory stimulus, then the visual recall prompt appeared until key press. Each visual stimulus was presented in Courier font 50-point size in white against a grey background. Participants were instructed to respond as quickly and accurately as they could to every item on the screen by deciding whether they heard that word or not, using the 'y' and 'n' keys on their keyboard. Participants had up to 3000 msec to make a decision. Reaction times were measured from the beginning of the visual recall prompt until key press, and the first key press was recorded.

4.4. Data analysis

The data, analysis code, and materials to reproduce the analyses and figures are available on OSF. https://osf.io/uv4ef/overview?view_only=1a410d6328c040a29186a3db01286f99.

All analyses were performed using the R Statistical Software (R Core Team, 2015), the 'lme4' package (Bates et al., 2015), and the 'DHARMA' package (Hartig, 2024). All files were cleaned using the same procedure: any data points beyond two standard deviations from participant means were counted as outliers and removed; participants with less than 70% accuracy were also removed.² Accuracy and reaction times (RTs) were analysed separately. Accuracy is a categorical outcome and was modelled using logistic mixed-effects models, whereas RTs are continuous and were analysed with linear mixed-effects models. Including error trials in RT analyses is inappropriate, as incorrect responses often reflect different processing states and exhibit distinct RT distributions, potentially distorting model estimates. Analysing the two measures separately therefore ensures appropriate model assumptions and clearer interpretation of effects.

² This threshold was determined a priori, based on common thresholds used in psycholinguistic experiments and screening practices for inattentive or disengaged participants (cf. Zwaan et al., 2018).

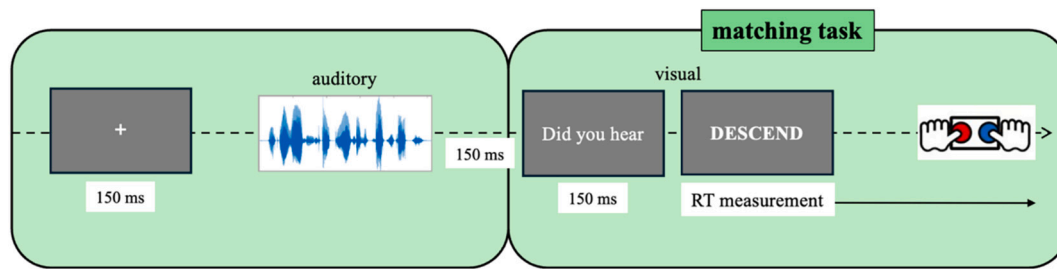


Fig. 1. Experimental design.

For the reaction times, statistical analyses were performed by fitting linear mixed-effects models using restricted maximum likelihood (REML). Participants were treated as random factors. A number of diagnostics were carried out for each analysis to check goodness-of-fit of the model to the data: these included tests for normality of residuals, absence of collinearity, and independence. For normality of residuals, scatterplots, histograms, and quantile-quantile (QQ) plots were visually inspected with respect to the presence or absence of patterns of heteroscedasticity. When a fixed effect or an interaction with it was significant, this was followed up by a pairwise comparisons test using Tukey HSD, so that the significance level applied jointly to all comparisons. For the error analysis, a binomial generalized linear model was used.

5. Experiment 1: expert processing

5.1. Participants

55 native speakers of British English were recruited for the experiment via social media message boards and word-of-mouth (mean age = 37.4). All participants reported speaking British English as their native language. 75% of participants reported having over 500 h of hour flown, with an additional 9% reporting over 2000 h. 85% of respondents were currently employed as flight crew (with 11% otherwise employed in an aviation industry) and 89% reported being based in the United Kingdom and/or Europe. The gender breakdown for respondents was 75.4% male/ 24.6% female. All participants gave their informed consent and were remunerated -appropriately for each 20-minute experiment. The experiment was overseen by the first author's university ethics review board.

5.2. Results

13 participants were excluded due to excessive response times or technical errors (N = 42). Data cleaning procedures resulted in an exclusion of an additional 14.5% of data. Reaction times (in ms, see Table 3) were analysed as a function of MATCH (match vs. mismatch) and

Table 3 Mean RTs (ms), SDs, and 95% CIs by match × legality for experts.

Auditory condition	Example	Recall condition	Recall word	Mean RT (ms)	SD RT (ms)	95% CI
Legal	Remain this frequency	match	REMAIN	1700	938	[1700 ± 51]
Legal	Remain this frequency	mismatch	STAY	1763	985	[1763 ± 48]
Illegal	Stay on this frequency	match	STAY	1688	916	[1688 ± 44]
Illegal	Stay on this frequency	mismatch	REMAIN	1744	954	[1744 ± 46]

LEGALITY (legal vs. illegal). The initial model for Reaction Times indicated significant main effects of MATCH, but no main effect of MATCH and no MATCH × LEGALITY interaction. Because the interaction was not significant, a reduced model excluding the interaction was refit. In the reduced model, MATCH remained a significant predictor, $\beta = 87.25, SE = 26.51, t = 3.29$, with mismatch trials producing longer RTs (see Table 4). LEGALITY did not significantly predict RTs, $\beta = -3.66, SE = 24.68, t = -0.148$. Thus, RTs were driven by whether the MATCH task contained a matching word to the utterance, not by whether the utterance was standard or non-standard.

Accuracy (as measured by error rates, see Table 5) was analysed using a logistic mixed-effects model³ with random intercepts by subject. There were significant main effects of MATCH ($\chi^2(1) = 41.83, p < .001^*$) and LEGALITY ($\chi^2(1) = 65.83, p < .001^*$), as well as a significant MATCH × LEGALITY interaction ($\chi^2(1) = 20.33, p < .001^*$) (see Tables 6 and 7). Predicted probabilities from the model indicated that accuracy was highest when the auditory condition contained Standard Aviation English and the recall word matched (e.g. *Remain this frequency* – *REMAIN*); lowest accuracy occurred when the utterance was non-standard and recall word matched (e.g. *Stay on this frequency* – *STAY*). Model diagnostics showed no violations of uniformity, dispersion, or zero inflation.

6. Experiment 2: non-expert processing

6.1. Participants

55 native speakers of British English were recruited for the experiment (mean age = 32.4) using the recruitment platform Prolific. All participants reported speaking British English as their native language; 3 participants additionally reported speaking a second language. No participants reported any familiarity with Aviation English. All participants gave their informed consent and were remunerated £10 for each 20-minute session. The experiment was overseen by the first author's university ethics review board.

6.2. Results

1 participant was excluded due to excessive RTs (N = 54). Data cleaning procedures resulted in an exclusion of an additional 11.1% of data. A linear mixed-effects model⁴ was fit to log-transformed reaction

Table 4 Fixed effects for reaction times for experts.

	Est	SE	t-Value
Intercept	1717.04	57.46	29.87
Match	87.25	26.51	3.291
Legality	-3.658	24.687	-0.148

³ glm(formula = Correct ~ match + legality, family = binomial, data = d).

⁴ lmer(logRT.x.1000 ~ match*legality+(1|date) + (1|Question), data = d).

Table 5
Error rates for experts.

Auditory condition	Example	Recall word	Recall condition	Error rate
Legal	<u>Remain</u> this frequency	REMAIN	match	6.13
Legal	<u>Remain</u> this frequency	STAY	mismatch	9.1
Illegal	<u>Stay on</u> this frequency	STAY	match	15.7
Illegal	<u>Stay on</u> this frequency	REMAIN	mismatch	8.3

Table 6
Analysis of deviance table (type III Wald chi-square tests) for errors for experts.

	Chisq	Df	Pr (>Chisq)
Intercept	1	412.48	p < .0001*
Match	1	41.83	p < .0001*
Legality	1	65.84	p < .0001*
Match*legality	1	20.33	p < .0001*

Table 7
Pairwise contrasts for error rates for experts.

Contrast	Estimate	SE	z-Ratio
match illegal - mismatch illegal	0.73	0.11	6.46
match illegal - match legal	1.08	0.13	8.11
match illegal - mismatch legal	0.96	0.12	7.95
mismatch illegal - match legal	0.355	0.15	0.07
mismatch illegal - mismatch legal	0.23	0.13	1.72
match legal - mismatch legal	-0.12	0.15	-0.81

times with MATCH, LEGALITY, and their interaction as fixed effects. Values for reaction times (see Table 8) were log-transformed values due to distinct skewing in the visual inspection of residuals: profile log-likelihoods were generated via a Box-Cox transformation ($\lambda = -0.58$). The analysis revealed no main effect of MATCH, $\beta = 0.0004$, $SE = 0.021$, $t = 0.02$, indicating that matched and mismatched trials did not differ in overall response speed (see Table 9). In contrast, LEGALITY showed a significant main effect, $\beta = 0.0447$, $SE = 0.0116$, $t = 3.87$, such that legal items elicited slower responses. Importantly, this effect was qualified by a significant MATCH \times LEGALITY interaction, $\beta = -0.0622$, $SE = 0.0269$, $t = -2.31$.

Pairwise comparisons (see Table 10) were conducted to clarify the interaction between MATCH and LEGALITY in the log-transformed reaction-time model. When comparing match-illegal to match-legal trials, responses were significantly slower for legal items, $\beta = -0.0447$, $SE = 0.0116$, $z = -3.87$. No difference emerged between match-illegal and mismatch-legal trials, $\beta = 0.0171$, $SE = 0.0260$, $z = 0.66$. Within

Table 8
Mean RTs (ms), SDs, and 95% CIs by match \times legality for non-experts.

Auditory condition	Example	Recall word	Recall condition	Mean RT (ms)	SD RT (ms)	95% CI
Legal	<u>Remain</u> this frequency	REMAIN	match	1022	414	[1022 \pm 22.5]
Legal	<u>Remain</u> this frequency	STAY	mismatch	1000	383	[1000 \pm 24.0]
Illegal	<u>Stay on</u> this frequency	STAY	match	934	375	[934 \pm 22.5]
Illegal	<u>Stay on</u> this frequency	REMAIN	mismatch	1075	410	[1075 \pm 24.5]

Table 9
Fixed effects for reaction times for non-experts.

Contrast	Estimate	SE	t-Ratio
Intercept	6.924	0.03803	182.1
Match	0.0004	0.02135	0.018
Legality	0.0447	0.01156	3.868
Match*legality	-0.0622	0.02686	-2.31

Table 10
Pairwise contrasts for reaction times for non-experts.

Contrast	Estimate	SE	z-Ratio
match illegal - mismatch illegal	-0.000374	0.0213	-0.018
match illegal - match legal	-0.044695	0.0116	-3.868
match illegal - mismatch legal	0.017129	0.0260	0.660
mismatch illegal - match legal	-0.044321	0.0217	-2.044
mismatch illegal - mismatch legal	0.017503	0.0243	0.719
match legal - mismatch legal	0.061824	0.0261	2.365

mismatch trials, mismatch-illegal responses were marginally faster than match-legal responses, $\beta = -0.0443$, $SE = 0.0217$, $z = -2.04$ (uncorrected). No significant difference was observed between mismatch-illegal and mismatch-legal trials, $\beta = 0.0175$, $SE = 0.0243$, $z = 0.72$. Finally, match-legal trials were slower than mismatch-legal trials, $\beta = 0.0618$, $SE = 0.0261$, $z = 2.37$. Overall, the pairwise comparisons reinforce the interpretation of the interaction: legality slowed responses, but this effect is stronger in match trials and largely absent in mismatch trials.

Recall that the utterances containing illegal (Violation) language contained more 'natural', linguistically well-formed constructions, e.g. *Stay on this frequency* vs. *Remain this frequency*. Non-experts appeared to find this 'expert language' more challenging to comprehend, resulting in larger RT differences for these items.

A logistic mixed-effects model⁵ was used to examine accuracy (as defined by error rates, see Table 11) as a function of MATCH and LEGALITY. The logistic mixed-effects model revealed significant effects of MATCH and LEGALITY on accuracy (see Table 12). Accuracy was lower on match trials relative to mismatch trials ($\beta = -0.86$, $SE = 0.11$, $z = -7.96$), indicating that participants were substantially less accurate when the recall word matched the auditory stimulus. LEGALITY also significantly affected accuracy ($\beta = -0.24$, $SE = 0.10$, $z = -2.36$), with legal items eliciting lower accuracy than illegal ones. There was no interaction ($\beta = 0.008$, $SE = 0.216$, $z = 0.041$). These effects suggest that participants were generally more accurate when the auditory stimulus and recall word were not aligned, and they performed better on non-standard (illegal) phrasing than on standard Aviation English phrasing.

Table 11
Error rates for non-experts.

Auditory condition	Example	Recall word	Recall condition	Error rate
Legal	<u>Remain</u> this frequency	REMAIN	match	13.5
Legal	<u>Remain</u> this frequency	STAY	mismatch	6.3
Illegal	<u>Stay on</u> this frequency	STAY	match	16.5
Illegal	<u>Stay on</u> this frequency	REMAIN	mismatch	7.9

⁵ glm(formula = Correct ~ match + legality, family = binomial, data = d).

Table 12
Fixed effects: Error rates for non-experts.

Predictor	Estimate	Std. error	z
Intercept	-1.7701	0.1189	-14.88
Match (mismatch)	-0.8563	0.1076	-7.96
Legality (legal)	-0.2429	0.1029	-2.36

7. Group analysis

For the group-level analysis, we modelled response times as a function of GROUP, MATCH and LEGALITY. A linear mixed-effects model⁶ was fitted for reaction times, with participant included as a random effect (see Table 13). There was a significant effect of GROUP on the match-mismatch contrast ($\beta = -22.5$, SE = 6.8, $z = -3.3$). Experts showed a reliable match effect, responding significantly faster on match than mismatch trials ($\beta = -76.62$, SE = 17.3, $z = -4.41$). In contrast, non-experts showed no meaningful match effect ($\beta = 13.53$, SE = 21.3, $z = 0.64$), indicating that any task-related RT differences present for experts were essentially absent in the non-expert group (see Table 14).

Accuracy was analysed using a binomial generalized linear mixed-effects model⁷ with fixed effects for GROUP, MATCH and LEGALITY, and a random intercept for participant (see Table 15). The model was fitted using *glmer* with the *bobyqa* optimizer and a maximum of 2×10^5 function evaluations. There was a three-way interaction between GROUP, MATCH and LEGALITY ($F(1, 10339) = 7.78$, $p = .005$): overall, experts showed stronger and more consistent effects of both LEGALITY and MATCH status on accuracy rates than non-experts (see Table 15). For experts, nonstandard utterances with MATCH (e.g. *Stay on this frequency- STAY*) trials produced reliably lower log-odds of correct responses than all other conditions, with large and significant contrasts against *illegal-mismatch*, *legal-match*, and *legal-mismatch* conditions (all $ps < .0001^*$, see Appendix A for comprehensive pairwise comparisons). Non-experts, by contrast, displayed a much flatter pattern, with only modest differences across legal and illegal items and smaller MATCH effects overall. Significant effects in the non-expert group were mostly confined to comparisons between match and mismatch conditions within the same LEGALITY category, but these effects were weaker than those observed for experts. Together, these findings suggest that experts were especially sensitive to non-standard language, and non-experts show comparatively diminished discrimination across conditions (Fig. 2).

8. General discussion

An expert's lexicon is a deeply interconnected, hierarchically and conceptually organized network of terms that supports rapid retrieval, problem-solving, and flexible use, reflecting both declarative knowledge and procedural expertise. The structure of expert language (hierarchies, categories, abbreviations) reflects the way that knowledge is organized

Table 13
Fixed effects reaction times for group comparison.

Term	Estimate	Std error	t ratio
Intercept	1420.0782	36.89567	38.49
Group	331.19483	36.89567	8.98
Legality	-4.850315	6.876798	-0.71
Group*legality	-1.563469	6.876798	-0.23
Match	-15.71533	6.878606	-2.28
Group*match	-22.69237	6.878606	-3.30
Legality*match	-1.147449	6.869562	-0.17
Group*legality*match	4.1129863	6.869562	0.60

⁶ *lmer*(logRT.x.1000 ~ match*legality*Group+(1|participant), data = d).

⁷ *glmer*(Correct ~match*legality*Group + (1|participant), family = binomial, data = d).

Table 14
Pairwise comparisons for reaction times for group comparison.

Group	match	-group	-match	Difference	Std error	t ratio
expert	match	expert	mismatch	-76.815	17.40482	-4.41
expert	match	non	match	617.005	75.10606	8.22
expert	match	non	mismatch	630.959	75.12502	8.4
expert	mismatch	non	match	693.82	75.0005	9.25
expert	mismatch	non	mismatch	707.774	75.01949	9.43
non	match	non	mismatch	13.954	21.30999	0.65

Table 15
Fixed effects for accuracy for group comparison.

Term	Estimate	Std error	DFDen	t ratio
Intercept	-2.800435	0.1538658	1164.5	-18.2
Group	0.0835391	0.1929177	1121.9	0.43
Legality	0.2224907	0.176531	10,339	1.26
Group*legality	0.0332939	0.2221283	10,339	0.15
Match	0.84103	0.1633108	10,339	5.15
Group*match	-0.940393	0.2237055	10,339	-4.2
Legality*match	0.0194486	0.2173972	10,339	0.09
Group*legality*match	0.8061595	0.2889678	10,339	2.79

cognitively. Expert knowledge and language are tightly linked: expert language encodes, builds, and activates deep conceptual understanding while enabling fast, precise communication and reasoning. While existing work has demonstrated that experts maintain distinct but overlapping lexical and conceptual representations for domain-specific and common-language terms, much of the empirical evidence comes from tasks involving complex reasoning, domain problem-solving, or multimodal processing.

In order to probe the organisation of domain-specific language in the expert mind, we used a sentence processing task combined with word recall to present experts (pilots) and non-experts (individuals with no knowledge of aviation radiotelephony) with auditory stimuli which either contained standard aviation phraseology or a nonstandard construction.

Our findings are as follows:

1. Accuracy patterns diverge sharply between experts and non-experts. Experts were highly sensitive to both legality and match word status, exhibiting large accuracy differences across conditions. For experts, standard aviation phraseology boosted performance: both types of recall items showed high accuracy ratings in the standard phraseology condition. Conversely, when the experts heard non-standard phraseology, errors increased, particularly when they heard then saw non-standard phraseology (e.g. *Stay on this frequency- STAY*).
2. Non-experts showed higher error rates on matched than mismatched trials in the within-group analysis, indicating difficulty in accurately recognising the auditory input. While experts have strong expectations for aviation phraseology, non-experts do not: for these listeners, *Remain this frequency* does not carry any special status. And although *Stay on this frequency* does not violate specific entries in the non-expert mental lexicon, both legal and illegal items are equally unfamiliar due to the special nature of the expert language. This suggests that, without aviation-specific schemas, non-experts form low-fidelity representations of the message, making recognition-based decisions (matches) harder than simple rejection decisions (mismatches). Furthermore, in the group analysis, these effects largely disappeared; non-experts exhibited a much flatter accuracy pattern than experts, with only minimal sensitivity to match status or legality.
3. Expertise confers large performance advantages, especially for illegal (nonstandard) items. Experts consistently outperformed non-experts on contrasts involving standard phraseology, with odds ratios indicating 2-3 times greater accuracy in critical comparisons. These

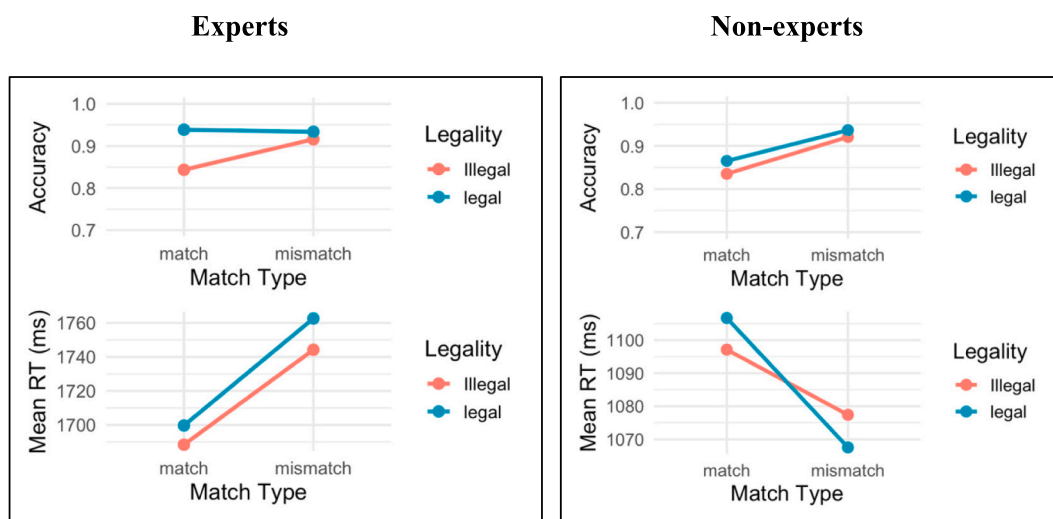


Fig. 2. Experts and non-experts comparisons.

effects were robust and widespread for experts but minimal for non-experts, suggesting that domain knowledge enhances sensitivity to both structural regularity and correct cue–response pairing.

- Experts show strong match effects in reaction times; non-experts do not. Experts responded much faster on match than mismatch trials, indicating efficient use of cue–response alignment. Non-experts showed no reliable RT difference, suggesting they did not benefit from the match relationship in the same way.

Taken together, our findings suggest that aviation language expertise is highly domain-specific, developing through extensive practice, high-quality feedback, and prolonged immersion in radiotelephony conventions. Experts do not simply possess more knowledge; they possess better structured and more proceduralized knowledge that allows them to rapidly integrate linguistic cues with operational expectations. This enables experts to detect illegal or unusual phraseology (e.g. *Stay on this frequency*), prioritise task-relevant information, and align auditory input with expected recall responses more efficiently and accurately than non-experts. In contrast, non-experts (who lack this specialised representational structure) show weaker sensitivity to incorrect/nonstandard phraseology and less reliable use of match cues.

In aviation, experts build up a mental vocabulary that is strongly tied to the systems they use, the procedures they follow, and the decisions they have to make. As a result, the phrases they encounter function not merely as linguistic input, but as cues that map onto well-practised actions, allowing for rapid and efficient processing. Achieving this level of expertise involves acquiring specialised linguistic patterns that operate across multiple layers of language, from phonology to phrasal structures. At the same time, however, an individual's general language system remains active. This broader linguistic knowledge, established through lifelong use, can introduce interference when everyday phrasing competes with the rigid and standardized forms required in aviation communication.

A growing body of work suggests that expert and non-experts often maintain parallel lexical representations for the same referent, with specialised and common lexical nodes coexisting (rather than replacing one another). This is often described as a distinction between the *general lexicon* and the *specialised lexicon*, each encoding different levels of conceptual granularity and participating in differently structured semantic networks (Cabré, 2010; Meyer & Mackintosh, 2000). Specialised lexical units tend to be embedded in dense, domain-specific relational structures that support precise inferencing, whereas their common-language counterparts reflect broader, sometimes naïve, conceptualizations of the same reality. Recent modelling work demonstrates that

these two lexica only partially overlap: experts develop specialised nodes with narrower meanings and stronger associative connectivity, while retaining access to the more general nodes shared with lay speakers (cf. Barbero & Amaro, 2024).

Our results support this dual-node account, showing that depth of lexical representation shapes how efficiently domain-specific words are accessed and how they compete with general-language alternatives. The interaction between these two systems (specialised and general) likely contributes to the performance differences observed between experts and non-experts, particularly when communications deviate from prescribed standards. Together, our results illustrate that expertise in aviation communication reflects not just familiarity with phraseology, but the development of an organized cognitive system optimised for interpreting, predicting, and responding to domain-specific language under time pressure.

Technological advances in psycholinguistic and neurolinguistic methodologies (e.g. online behavioural experiments, eye-tracking, and brain imaging) are becoming increasingly accessible, laying the pathway for more effective research collaborations. Our findings contribute important psycholinguistic evidence that specialised language systems are mentally represented in structured ways: simply put, expertise fundamentally reshapes language processing. They also provide new evidence about the cognitive role of domain-specific constraints, suggesting that expertise leads to strong predictive expectations about allowable forms. Last (but not least), the findings also have practical relevance for the field of aviation communication. They demonstrate that standard phraseology is not merely procedural, but cognitively beneficial. Deviations to standard phraseology can increase misunderstanding even for trained professionals.

9. Limitations to the study and future directions

A key limitation of this study concerns the online mode of data collection. Although online experiments allow access to broader participant samples in some cases, measurements collected in uncontrolled environments can be less precise than those obtained in laboratory settings. Variability in participants' hardware, software, internet latency, and input devices may introduce additional noise into responses (particularly with reaction time measurements), potentially reducing reliability.

A second limitation to the study stems from its reliance exclusively on behavioural measures. Although these measures contribute valuable insights into processing efficiency and decision-making, they do not directly reveal neural mechanisms involved in the processing of domain-

specific phraseology (and specifically, deviations to that phraseology). Behavioural outcomes represent the end product of multiple cognitive processes, making it difficult to determine *when* deviations begin to affect comprehension. Thus, future research could therefore extend this work using neurophysiological methods such as electroencephalography (EEG), allowing us to elicit neural signatures associated with linguistic predictions (N400) or increased integration effort (P600), providing a more detailed account of how expertise shapes language processing in expert domains.

CRedit authorship contribution statement

Hilary Wynne: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Henry Emery:** Resources, Methodology, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, we used ChatGPT 5.0 to assist with editing R scripts written and prepared by the first author. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

We have nothing to declare.

Acknowledgements

This publication arises from research funded by the John Fell Oxford University Press Research Fund awarded to Dr. Hilary Wynne.

Appendix A. Pairwise comparison by group

Group	Legality	Match	-Group	-Legality	-Match	Estimate	Std error	t Ratio
expert	Illegal	match	expert	Illegal	mismatch	0.726245	0.11341933	6.40
expert	Illegal	match	expert	legal	match	1.081393	0.13443460	8.04
expert	Illegal	match	expert	legal	mismatch	0.982030	0.12148835	8.08
expert	Illegal	match	non	Illegal	match	-0.017400	0.14547834	-0.12
expert	Illegal	match	non	Illegal	mismatch	0.843078	0.16552503	5.09
expert	Illegal	match	non	legal	match	0.224539	0.15094385	1.49
expert	Illegal	match	non	legal	mismatch	1.065569	0.17837191	5.97
expert	Illegal	mismatch	expert	legal	match	0.355147	0.14658725	2.42
expert	Illegal	mismatch	expert	legal	mismatch	0.255785	0.13482502	1.90
expert	Illegal	mismatch	non	Illegal	match	-0.743646	0.15703851	-4.74
expert	Illegal	mismatch	non	Illegal	mismatch	0.116833	0.17577168	0.66
expert	Illegal	mismatch	non	legal	match	-0.501706	0.16211476	-3.09
expert	Illegal	mismatch	non	legal	mismatch	0.339324	0.18791936	1.81
expert	legal	match	expert	legal	mismatch	-0.099363	0.15288471	-0.65
expert	legal	match	non	Illegal	match	-1.098793	0.17292513	-6.35
expert	legal	match	non	Illegal	mismatch	-0.238314	0.19009916	-1.25
expert	legal	match	non	legal	match	-0.856854	0.17754774	-4.83
expert	legal	match	non	legal	mismatch	-0.015824	0.20138444	-0.08
expert	legal	mismatch	non	Illegal	match	-0.999430	0.16298662	-6.13
expert	legal	mismatch	non	Illegal	mismatch	-0.138952	0.18110557	-0.77
expert	legal	mismatch	non	legal	match	-0.757491	0.16788311	-4.51
expert	legal	mismatch	non	legal	mismatch	0.083539	0.19291768	0.43
non	Illegal	match	non	Illegal	mismatch	0.860479	0.14387108	5.98
non	Illegal	match	non	legal	match	0.241939	0.12709388	1.90
non	Illegal	match	non	legal	mismatch	1.082969	0.15861994	6.83
non	Illegal	mismatch	non	legal	match	-0.618539	0.14920970	-4.15
non	Illegal	mismatch	non	legal	mismatch	0.222491	0.17653097	1.26
non	legal	match	non	legal	mismatch	0.841030	0.16331081	5.15

Appendix B. Stimuli

Standard phraseology

Item	Standard phraseology	Match	Mismatch
1	Taxi to holding point Alpha one runway two two.	RUNWAY 22	EXPEDITE
2	PAPI runway zero niner unserviceable.	UNSERVICEABLE	RETURN
3	When vacated contact ground one two one decimal niner.	VACATED	TOWER
4	Descend flight level one hundred. Speed at your discretion.	DISCRETION	ASCEND
5	Cleared direct LORIX.	DIRECT	LOWER
6	ILS approach not available due equipment failure.	DUE	AVAILABLE
7	Cleared to Oscar Whisky via MERIN one Echo arrival.	VIA	HERON
8	Maintain FL310 until advised by Wembury control.	ADVISED	REMAIN
9	Report runway in sight.	SIGHT	TAXIWAY
10	Resume own navigation one one miles south of Alpha November November.	RESUME	NORTH
11	Confirm RVSM approved.	WHEN	AREA
12	Contact Wembury Control 124 decimal 175	CONTACT	ROUTINE

(continued on next page)

(continued)

Item	Standard phraseology	Match	Mismatch
13	Maintain own separation.	MAINTAIN	CLEARED
14	ILS category two not available.	UNABLE	ACCEPT
15	Increase speed two three zero knots.	INCREASE	NOVEMBER
16	Cross SAMAK at FL 310.	CROSS	RUNWAY
17	Take first left and contact ground one two one decimal six.	TAKE	DELTA
18	Do not exceed three zero zero knots.	EXCEED	LOWER
19	Expect further clearance at time five one.	EXPECT	LIMIT
20	Go around immediately.	GO AROUND	UPPER
21	Are you ready for pushback?	READY	PRESS
22	Runway zero niner , wind one hundred degrees seven knots, QNH one zero one two.	WIND 100 DEGREES	KNOTS
23	Turn left heading zero two zero to FL70 to intercept upper Mike eight five one.	INTERCEPT	RIGHT
24	Report established on the localiser.	ESTABLISHED	VECTOR
25	Fly heading zero six zero. When able proceed direct SAMAK.	ABLE	CHANNEL
26	Remain this frequency	REMAIN	POSITION
27	Descend flight level one two zero.	DESCEND	CONFIRM
28	Maintain FL180.	MAINTAIN	GROUND
29	Reduce speed by two zero knots.	SPEED	DISTANCE
30	Report Oscar Whisky outbound.	REPORT	FINAL
31	Advise able to accept visual approach runway zero niner.	ADVISE	MARKER
32	After passing five thousand feet turn right direct MYN.	PASSING	ZAMAK
33	Caution work in progress at entry to apron.	CAUTION	RIGHT
34	Cleared for take-off from intersection Bravo runway one three.	CLEARED	ADVISED
35	Taxi to holding point alpha one runway three four.	TAXI	TWENTY
36	Climb flight level three two zero.	CLIMB	TAKE
37	Cancel offset, Return back to upper Mike seven three zero via MERIN.	CANCEL	LEFT
38	Enter controlled airspace via MERIN at FL150 at time four seven.	ENTER	TAXIWAY
39	Join upper golf 1 at MERIN at FL330 at time three two.	JOIN	HERON
40	Leave controlled airspace via MERIN at FL110.	LEAVE	LIMA

Nonstandard phraseology (violation in bold)

Item	Non-standard phraseology	Match	Mismatch
1	Taxi to holding point alpha one runway twenty two .	RUNWAY 22	READY
2	PAPI runway zero niner isn't working .	WORKING	WIND
3	When you've left the runway contact ground one two one decimal niner.	LEFT	INTERCEPT
4	Descend flight level one hundred. Speed as you want .	YOU WANT	ESTABLISHED
5	Cleared straight to LORIX.	STRAIGHT	ABLE
6	ILS approach not available because of equipment failure.	BECAUSE	REMAIN
7	Cleared to Oscar Whisky taking the MERIN one Echo arrival.	TAKING	DESCEND
8	Maintain FL310 until Wembury control tell you otherwise .	TELL YOU	MAINTAIN
9	Report visual contact with runway.	CONTACT	SPEED
10	Go back to own navigation one one miles south of ANN.	GO BACK	REPORT
11	Verify if you're RVSM approved	IF YOU'RE	ADVISE
12	Get in touch with Wembury Control 124 decimal 175	GET IN TOUCH	PASSING
13	Keep own separation.	KEEP	CAUTION
14	We can't give you ILS category two.	CAN'T	CLEARED
15	Accelerate to speed two three zero knots.	ACCELERATE	TAXI
16	Fly over SAMAK at FL 310.	OVER	CLIMB
17	Hang first left and contact ground one two one decimal six.	HANG	CANCEL
18	Don't go faster than three zero zero knots.	FASTER	ENTER
19	Anticipate further clearance at time five one.	ANTICIPATE	JOIN
20	Abort landing immediately	ABORT	LEAVE
21	Are you good-to-go for pushback?	GOOD	RUNWAY 32
22	Runway niner, wind one hundred degrees seven knots, QNH one thousand and two .	WIND 100 DEGREES	UNSERVICEABLE
23	Turn left heading zero two zero to FL70 to get onto upper Mike eight five one.	TO GET ONTO	VACATED
24	Report when you're on the localiser.	WHEN	DISCRETION
25	Fly heading zero six zero. When you can proceed direct SAMAK.	CAN	DIRECT
26	Stay on this frequency.	STAY	DUE
27	Go down to flight level one two zero.	DOWN	VIA
28	Stay at to FL180.	STAY	ADVISED
29	Slow down by two zero knots.	SLOW DOWN	SIGHT
30	Tell us when you're Oscar Whisky outbound.	WHEN	RESUME
31	Tell me if able to accept visual approach runway zero niner.	TELL ME	WHEN
32	Once above five thousand feet turn right direct Mike Yankee November.	ABOVE	CONTACT
33	Be careful . Work in progress at entry to apron.	CAREFUL	MAINTAIN
34	Permitted to take-off from intersection Bravo runway one three.	PERMITTED	UNABLE
35	Move to holding point alpha one runway three four.	MOVE	INCREASE
36	Go up to flight level three two zero.	UP TO	CROSS
37	Abandon offset, double back to upper Mike seven three zero via MERIN.	ABANDON	TAKE
38	Come in to controlled airspace via MERIN at FL150 at time four seven.	COME IN	EXCEED
39	Get on to upper golf 1 at MERIN at FL330 at time three two.	GET ON	EXPECT
40	Depart controlled airspace via MERIN at FL110.	DEPART	GO AROUND

Data availability

OSF

References

- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 10(3), 483.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons.
- Ballot, C., Robert, C., Dujardin, E., & Mathey, S. (2024). Effects of lexical skills and orthographic neighborhood size in word memory. *Memory & Cognition*, 52(3), 610–621.
- Barbero, C., & Amaro, R. (2024). Are we talking about the same thing? Modeling semantic similarity between common and specialized lexica in WordNet. *Languages*, 9(3), 89.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bergum, A., Maurer, A. M., Peitek, N., Bader, R., Mecklinger, A., Demberg, V., & Apel, S. (2024). Unexpected but informative: What fixation-related potentials tell us about the processing of confusing program code. *arXiv preprint*, Article arXiv:2412.10099.
- Bilenko, N. Y., Grindrod, C. M., Myers, E. B., & Blumstein, S. E. (2008). Neural correlates of semantic competition during processing of ambiguous words. *Journal of Cognitive Neuroscience*, 21(5), 960–975.
- Bryce, T. G. K., & Blown, E. J. (2012). The novice-expert continuum in astronomy knowledge. *International Journal of Science Education*, 34(4), 545–587.
- Cabr , M. T. (2010). Terminology and translation. *Handbook of Translation Studies*, 1, 356–365.
- Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving*. No. TR5.
- De Groot, A. D. (1965). *Thought and choice in chess*. Amsterdam University Press.
- Firth, J. R. (1957). Applications of general linguistics. *Transactions of the Philological Society*, 56(1), 1–14.
- Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition*, 66(2), 115–152.
- Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, 6(3), 225–255.
- Gotti, M. (2008). *Investigating specialized discourse*. Peter Lang.
- Harris, Z. (1982). Discourse and sublanguage. In R. Kittredge, & J. Lehrberger (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 231–250). Berlin: de Gruyter.
- Hartig, F. (2024). DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.7. <https://CRAN.R-project.org/package=DHARMA>.
- Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In *Exploring expertise: Issues and perspectives* (pp. 81–100). London: Palgrave Macmillan UK.
- Kalverk mper, H. (1983). Textuelle Fachsprachen-Linguistik als Aufgabe/“Textual linguistics of languages for special purposes: A new field of research”. *Zeitschrift f r Literaturwissenschaft und Linguistik*, 13(51), 124.
- Kalverk mper, H. (1996). Im Zentrum der Interessen: Fachkommunikation als Leitgr e. *HERMES-Journal of Language and Communication in Business*, 16, 117–176.
- Kim, K., Bae, J., Nho, M. W., & Lee, C. H. (2011). How do experts and novices differ? Relation versus attribute and thinking versus feeling in language use. *Psychology of Aesthetics, Creativity, and the Arts*, 5(4), 379.
- Kittredge, R. (1982). Variety and homogeneity of sublanguages. In R. Kittredge, & J. Lehrberger (Eds.), *Sublanguage: Studies of languages in restricted domains*. (pp. 107–137). Berlin: de Gruyter.
- Kozma, R. B. (2020). Use of multiple representations by experts and novices. In *Handbook of learning from multiple representations and perspectives* (pp. 33–47). Routledge.
- Kuo, C. H., & Prat, C. S. (2024). Computer programmers show distinct, expertise-dependent brain responses to violations in form and meaning when reading code. *Scientific Reports*, 14(1), 5404.
- Masrai, A., Milton, J., El-Dakhs, D. A. S., & Elmenshawhy, H. (2021). Measuring the contribution of specialist vocabulary knowledge to academic achievement: Disentangling effects of multiple types of word knowledge. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 8.
- Maturi, K. S., & Sheridan, H. (2020). Expertise effects on attention and eye-movement control during visual search: Evidence from the domain of music reading. *Attention, Perception, & Psychophysics*, 82(5), 2201–2208.
- Meyer, I., & Mackintosh, K. (2000). When terms move into our everyday lives: An overview of de-terminologization. *International Journal of Theoretical and Applied Issues in Specialized Communication*, 6(1), 111–138.
- Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 144.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125(8), 1063–1068.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225–236.
- Postal, V. (2004). Expertise in cognitive psychology: Testing the hypothesis of long-term working memory in a study of soccer players. *Perceptual and Motor Skills*, 99(2), 403–420.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rikers, R. M., Schmidt, H. G., Boshuizen, H. P., Linssen, G. C., Wesseling, G., & Paas, F. G. (2002). The robustness of medical expertise: Clinical case processing by medical experts and subexperts. *The American Journal of Psychology*, 115(4), 609–629.
- Schiller, N. O. (2021). *The mental lexicon*. Oxford University Press.
- Schmidt, H. G., & Boshuizen, H. P. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205–221.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61(4), 394–403.
- Simon, H. A., & Gilmarin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5(1), 29–46.
- Staszewski, J. J. (1990). *Skill and skilled memory*. Final report. Research report 04206.
- Strevens, P., & Johnson, E. (1983). SEASPEAK: A project in applied linguistics, language engineering, and eventually ESP for sailors. *The ESP Journal*, 2(2), 123–129.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645–659.
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797.
- Toth, A. J., Kowal, M., & Campbell, M. J. (2019). The color-word Stroop task does not differentiate cognitive inhibition ability among esports gamers of varying expertise. *Frontiers in Psychology*, 10, 2852.
- Troyer, M., & Kutas, M. (2020). To catch a snitch: Brain potentials reveal variability in the functional organization of (fictional) world knowledge during reading. *Journal of Memory and Language*, 113, Article 104111.
- Vintar, S., & Saksida, A. (2023). The anatomy of specialized knowledge: Comparing experts and non-experts through associations, frames and language models. *Lexicographica*, 39(1), 165–190.
- Walla, P., Kalt, S., & Lachmayer, K. (2024). Neurophysiological correlates of expert knowledge: An event-related potential (ERP) study about law-relevant versus law-irrelevant terms. *Brain Sciences*, 14(10), 1029.
- Wallace, W. D. (1981). How registers register: Toward the analysis of language use. *IRAL: International Review of Applied Linguistics in Language Teaching*, 19(4), 267.
- Wiedenbeck, S. (1985). Novice/expert differences in programming skills. *International Journal of Man-Machine Studies*, 23(4), 383–390.
- Wynne, H. S. Z. (2025). The psychological reality of ESP processing: Insights from Aviation English. In *Selected papers of the 2025 APLX, ETA & TESPA International Conference on English Language Teaching and Learning* (Taipei, Taiwan).
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, Article e120.