Paper Type: Full paper/article: 12-to-20-pages

How the brain processes Aviation English:
Crafting a neurolinguistic perspective on the effects of nonstandard phraseology

HILARY WYNNE[a1] JIAMIN CHENG[b1] HENRY EMERY[c 1,2]
[1]University of Oxford, United Kingdom
[2]Latitude Aviation English Services, United Kingdom

Abstract
Over the past two decades, advances in neurolinguistic methodologies have enabled increasingly detailed investigation of the neural basis of language processing. Techniques such as EEG, MEG, and fMRI provide new insights into the psychological reality of language but also underscore the need for theoretically grounded frameworks to guide the interpretation and application of such data. This paper outlines recommendations for applying neurolinguistic evidence to the study of Aviation English, emphasising the importance of theory-informed experimental design. We present data from a project which examines how non-standard phraseology is processed in the expert brain. By situating neurolinguistic approaches within a broader linguistic and applied context, the paper aims to bridge the gap between experimental findings and their implications for language teaching, assessment, and professional communication.

Introduction

Over the past two decades, research into the neurological foundations of language has expanded considerably. Advances in methodologies and technologies now enable the collection and analysis of linguistic data at unprecedented speed and scale. Neurolinguistic techniques like EEG (electroencephalography), MEG (magnetoencephalography), and fMRI (functional magnetic resonance imaging) are regularly employed for researching brain activity related to and during linguistic processing, with the aim to provide insights into the

---

[a] Hilary Wynne is a senior researcher in the Language and Brain Laboratory at the University of Oxford. She has more than fifteen years of experience investigating linguistic processing using psycholinguistic and neurolinguistic methodologies and brain-imaging techniques.

[b] Jiamin Cheng is a research assistant at the Language and Brain Laboratory at the University of Oxford working on bilingual word production and expert English processing using EEG.

[c] Henry Emery is an aviation language education and assessment practitioner with a particular interest in instructional design and language test construction and evaluation. He is managing director of Latitude Aviation English Services, UK, a specialist provider of aviation language training and testing services and a board member of the International Civil Aviation English Association.

psychological reality of language. At the same time, this situation highlights the ongoing need for robust frameworks to guide interpretation and application of the data gained through these methods, for only by grounding experimental design firmly in relevant theoretical frameworks can we produce reliable results.

The goal of this paper is to outline recommendations for the application of neurolinguistic data to the study of Aviation English and demonstrate how theory-informed experimentation might operate in practice. We present a case study from a collaborative project (*ELSSE: Eliciting language-specific signatures for expertise in the human brain*) designed to investigate the effects of nonstandard phraseology on the expert brain. By situating neurolinguistic research within a broader theoretical and applied linguistic context, we seek to bridge the gap between empirical findings and their practical relevance for language teaching, assessment, and professional communication.

Standard and Nonstandard Phraseology

The issues revolving around standard phraseology in aviation are not merely pedagogical, but stem from a deeper misalignment between language as a system and language as a situated practice. As we know, the functional scope of a specialised language can be inherently limited, and once communicative demands exceed its predefined range, language usage often shifts. Regulatory guidance and testing regimes have traditionally focussed on abstract, decontextualised notions of ‘proficiency’ over the interactional, real-life nature of aeronautical communication, resulting in generic test designs that fail to capture pragmatic, discourse-oriented demands of radiotelephony. Crucially, there remains a lack of empirical evidence regarding the real-world effects of nonstandard phraseology on the listener: while it often assumed to degrade safety or clarity, there is no neurolinguistic research demonstrating how, when and to what extents deviations actually impact comprehension, repair or operational outcomes. This evidentiary gap further complicates pedagogy, as instructors and regulators are left to rely on intuition, anecdotes or prescriptive norms rather than robust neurocognitive data, reinforcing uncertainty about what should be taught, assessed or prioritised in practice (cf. Friginal et al., 2020).

Language, Expertise, and the Brain

When perceiving language, we hear words and phrases embedded within sentences: these linguistics structures cue our brains to access specific stored mental representations. According to theories of cue-based sentence comprehension (e.g. Van Dyke & McElree, 2011), when processing a sentence, a listener links incoming words and phrases with memory cues that point towards networks of linguistic constructions. Consequently, listeners are not only processing individual words but also forming predictions about what comes next, based on the context of the sentence. A number of prominent studies have focussed on the influence of contextual information on language processing (Kutas & Hillyard, 1984; Federmeier & Kutas, 1999): one reliable finding is that, when sentences contain highly informative contexts (as in Example 1), participants' responses are often facilitated (i.e. faster and more accurate), reflecting a boost in processing:

Example 1: *A female chicken is a ...*

In the above example, the auxiliary verb *is* cues a search for an appropriate subject NP (noun phrase) based on a combination of grammatical specifications and contextual cues learned through the accumulation of real-world knowledge (i.e. *female chicken* = *hen*$_{\text{NP}}$). Real-world knowledge is made up of an individual's semantic and lexical knowledge (Altmann & Kamide, 1999) which is continuously shaped by their own perceptual experience; consequently, individual differences can lead to significant variation in how we process language (cf. Novick et al., 2005; Troyer & Kutas, 2020) and sentences with highly informative contexts (Example 2) can have a number of possible predictable completions:

Example 2: *A female chicken is a...* ***hen*** */ bird/ fowl.*

However, a poultry farmer, whose lexicon would presumably contain more detailed entries for chickens than the average speaker, may be inclined to complete the sentence with *pullet* (a juvenile female chicken who is not yet laying eggs).

This raises questions not only about the nature of an expert's mental lexicon but also what may be involved in processing linguistic codes for a particular communicative purpose. A prominent feature of specialised languages such as Aviation English is **mono-referentiality**: defined as a quality by which words and phrases in a specialised language can only have a single referent (cf. Gotti, 2008). Thus, unlike Example 2 (where a number of synonymous words to *hen* can be used to complete the sentence), *maintain* in Aviation English cannot be replaced with any other word or phrase:[d]

Example 3: *Climb and* ***maintain*** */ *stay at / *go to / *remain at… 7000.*

The effects of expertise on perception and processing have been studied in expert domains such as computer coding (Adelson, 1984; Kuo & Prat, 2024), chess (de Groot, 1965; Chase & Simon, 1973; Simon & Gilmartin 1973; Gobet, 1998; Gobet & Simon, 2010), physics (Chi et al., 1982), medical expertise (Norman, Brooks, & Allen, 1989; Rikers et al., 2002; Schmidt & Boshuizen, 1993), astronomy (Bryce & Blown, 2012), and music (Maturi & Sheridan, 2020). Unsurprisingly, findings in the neurolinguistic literature (cf. Troyer & Kutas, 2020; Kuo & Prat, 2024) have shown that experts who learn a specialised language are specially cued to disruptions in the languages' structures. For example, Walla et al. (2024) found significant responses in legal experts' EEG signals for incongruent (law-irrelevant) stimuli in the legal

---

[d] Following standard linguistic notation, we use * to denote illegal constructions.

context, and in a study investigating the effect of training level on computer coding, Kuo & Prat (2024) found that expert coders exhibited strong sensitivities in linguistically relevant neural signals to semantic errors in code (e.g. *FOR FRUIT IN PETS: PRINT(FRUIT)*).

But violations come in many shapes and sizes: for example, violations to category expectations such as those in Kuo & Prat’s (2024) stimuli can cross a wide range of acceptability (compare, for example, *FOR FRUIT IN **PETS**: PRINT(FRUIT)* to a within-category violation in *FOR FRUIT IN **VEGETABLES**: PRINT(FRUIT)*). Violations to standard aviation phraseology also occur across a spectrum of what can be considered more or less nonstandard (e.g. ***keep** heading 050* vs. *he’s bust his level,* cf. McMillan, 1998; Krifka et al., 2003; Howard, 2008; Wynne, 2025). Accordingly, one of the questions our research aims to investigate is how different degrees of violation (i.e. nonstandard phraseology) affect the processing of language in the aviation environment.

Example of the project, preliminary findings

Design

This project delves into uncharted territory, both for the fields of aviation language research and neurolinguistics. Although a growing body of research has focused on differences between experts and non-experts, evidence for the explicit contribution of language to expertise is scarce. To investigate this, we planned a pilot (i.e. a ‘trial run’) experiment to test the effects of different types of violations of standard phraseology using event-related EEG, which measures electrical activity in the brain during real-time comprehension of (amongst other things) language. In research EEG, electrodes on the scalp record thousands of simultaneously ongoing brain processes while a participant receives stimuli (e.g. sounds or images): these processes can be translated into specific signals of interest, or event-related potentials (ERPs).

One type of established brain response is the N400, a negative-going brain response occurring between 300 and 500 milliseconds after a stimulus that reflects a violation of semantic expectations, e.g. *a female chicken is a SHOE* (cf. Hillyard & Kutas, 1984 and Figure 1 below).

Figure 1: the N400. The x-axis shows time (in ms) after the trigger, the y-axis shows differences in voltage.

We set out to elicit this response in aviation experts using a standard sentence processing task (auditory comprehension + *0*-back response, cf. Kirchner, 1958) combined with EEG. In a standard sentence processing paradigm, participants are asked to make judgements or complete a related task while listening to (or reading) the critical stimuli. We presented our participants with three types of auditory stimuli: standard (Example 4a), nonstandard (4b), and incongruent (i.e. impossible) (4c) utterances in Aviation English.

4a) *ILS category two not available.*
4b) *ILS category two not possible.*
4c) *ILS category two not dressing.*

Stimuli were normed for each category by means of an online version of the sentence processing task, in which pilots with more than 200 hours of flight time (N=43) provided behavioural measurements (reaction times and accuracy ratings) for utterances containing standard and non-standard phraseology (cf. Wynne & Emery, 2026). Recordings of the stimuli were made by a native speaker of standard British English familiar with the cadence, rate, and prosody of Aviation English. These recordings were made with low-level authentic tower chatter to replicate the standard operating environment. After recording, the experimenter checked all recordings, ensured they were all of a specific pitch and loudness (75 db, with an average signal-to-noise ratio of 22.2db), and measured the durations of each utterance.

Participants were told they would be doing a matching task: after hearing an auditory stimulus, the participant would see a word on a screen: the word either matched (e.g. *available*) or mismatched (e.g. *hospital*) to a word in the sentence (see Table 1). The participants were told that their job was to respond ‘yes’ or ‘no’, using a handheld game controller, whether they had heard the word in the auditory sentence or not. We chose this paradigm in order to not force a judgement response in terms of ‘correct’ or ‘incorrect’, (a common judgement task in sentence processing studies) as well as to redirect the participants’ attention on the stimulus, as the word they saw on the screen was often the critical word (standard or nonstandard) of the sentence.

Table 1: stimuli design

Utterance Type	Task Condition
STANDARD PHRASEOLOGY *ILS category two not available.*	**MATCH** *available*
	MISMATCH *hospital*
NONSTANDARD PHRASEOLOGY *ILS category two not possible.*	**MATCH** *possible*
	MISMATCH *occasion*
INCONGRUENT PHRASEOLOGY *ILS category two not dressing.*	**MATCH** *dressing*
	MISMATCH *quarter*

We presented our stimuli in the auditory modality, as this best reproduces the natural environment in which Aviation English is used. 24 standard, 24 nonstandard, and 24 illegal utterances were constructed, with special attention paid to the semantic distance of the critical items to one another, as defined by Latent Semantic Analysis (LSA). LSA (cf. Deerwester et al., 1990 and Landauer, Foltz & Laham, 1998) is a method for estimating semantic similarity by analysing patterns of word co-occurrence in a large text corpus. It builds a word-document matrix, reduces it with singular value decomposition (SVD), and represents words or texts as vectors in a shared semantic space. Items that appear in similar contexts end up close together, allowing researchers to quantify how semantically related different stimuli are. For sentence stimuli, LSA can evaluate semantic overlap between verbs and nouns, coherence across sentence context, and plausibility based on world knowledge patterns. Our standard (e.g. *available*) and nonstandard (*possible*) critical words had high similarity rates (mean cosine value of 0.287), and our illegal items (*dressing*) had very low similarity rates to both the standard and nonstandard items (~0.06). Auditory stimuli were coded for duration, word frequency, and number of phonological segments of the critical word was included in the pre-experimental coding; visual stimuli were coded for length (number of letters), and word frequency.

In psycholinguistic and neurolinguistic experiments, filler items are regularly included alongside the critical test items in order to prevent participants from becoming aware of the specific linguistic patterns or manipulations under investigation. Fillers also help maintain a natural processing environment by increasing variability, reducing predictability, and distributing experimental conditions more evenly. This improves the validity of the data by ensuring that participants’ responses to test items reflect typical language processing rather than task-induced expectations. We included an additional 72 filler items which contained either correct or incorrect phraseology: this resulted in 144 total trials in the experiment.

Conditions were counterbalanced across sessions such that participants saw an equal number of items in each session. Trials were also pseudorandomized in order to distance similar utterances from one another and to avoid repetition of the same utterance type within one session.

EEG requires hardware triggers to be defined prior to recording: these are time-locked signals sent to the recording system to mark precisely when an experimental event occurs so that neural activity can be aligned with that event during analysis. Due to the fact that they are implemented via the hardware (and not the software), these triggers are near-simultaneous with the event during the timescales of an EEG recording. Since our sentences were natural utterances and were all different durations, our triggers had to be hand-defined according to onset times of the critical word in the stimulus (e.g. *available/ possible/ dressing*) in milliseconds. Accordingly, we had 72 unique triggers for the test stimuli (fillers were coded with 4 trigger codes corresponding to filler condition). We included a second trigger at the onset of the visual target, both as a backup (in case the auditory trigger failed) and also to provide an additional measurement of neural responses to correct, incorrect, and illegal items during the matching task (see Figure 1 below for experimental timeline and trigger plan).

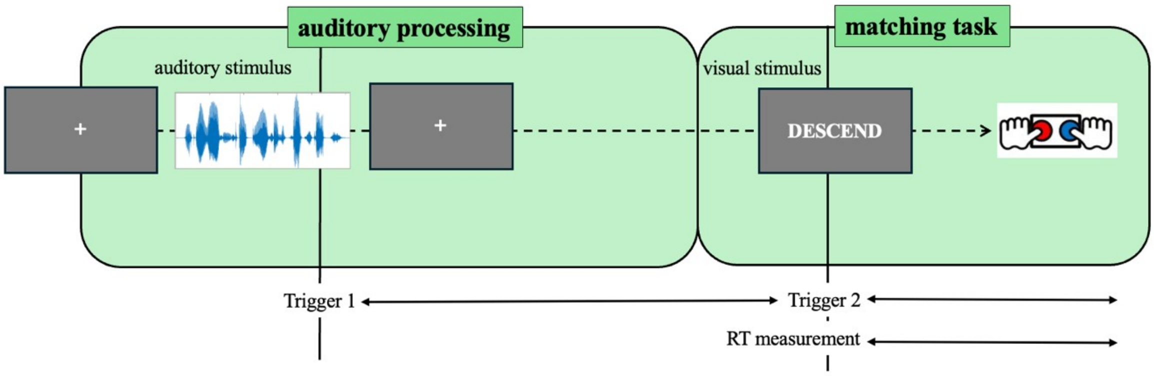

Figure 2:experimental timeline

Procedure

The experiment took place in the Language and Brain Laboratory at the University of Oxford, United Kingdom. During each session, participants were seated in a soundattenuated, shielded booth at a comfortable distance from a screen. The experiment was coded in PsychoPy (v2025.2.2, Pierce et al., 2019) and administered via Pavlovia (Open Science Tools, Nottingham, UK). Each trial began with a fixation cross for 150 ms, followed by the auditory stimulus, another fixation cross for 150 ms, then the visual recall prompt appeared until key press (Figure 1). Each visual stimulus was presented in Courier font 50-point size in white against a grey background. Reaction times (in ms) were measured from the beginning of the visual recall prompt until key press, and only the first key press was recorded. Participants first participated in a practice session consisting of 10 practice items in which feedback on correct responses was provided ('Correct/Incorrect'); there was no feedback in the main task. There were four blocks of 6 minutes each and participants were given the option of short breaks after every block. Including setup, the session lasted 90 minutes.

Participants

Our participants were recruited for the experiment via social media message boards, messaging services, and word-of-mouth (see Figure 3 for participant demographics). We chose to include both pilots and controllers in our expert population because, although the operational roles differ, both professions are trained to use and interpret the same specialised language during radiotelephony communication. Nevertheless, operational role was coded into the analysis to ensure there was no effect. All participants had normal or corrected-to-normal vision and gave their informed consent and were remunerated appropriately for their time and travel. No history of neurological disorders or hearing deficits was reported. The experiment was overseen by the University of Oxford's central ethics review board.

When venturing into new experimental territory- such as implementing a novel paradigm or introducing a new presentation modality- it is essential for experimenters to minimize potential confounds, including those arising from participants' language backgrounds. Thus, our recruitment procedures for this experiment were extremely strict: participants were required to be monolingual speakers of British English between the ages of 18 and 65, with over 250 hours of flight or duty time. In fact, the majority of our participants (75%) were highly skilled and had over 1000 hours of flight/duty time.

Figure 3: Participant demographics: ATC = air traffic controller, COMM= commercial pilot, MIL= military pilot, PPL = private pilot.

Measurement and Analysis

Continuous EEG is recorded with 32 active electrodes (ActiCap, Brain Products, Inc., Gilching, Germany). Electrodes are placed according to the international 10–10 system. During data acquisition, FCz is used as reference channel and the data is sampled at 1000 Hz. Impedance is kept below 25 kOhm wherever possible. Our EEG data was analysed using the EEGlab (Delorme & Makeig, 2004) and ERPlab (Lopez-Calderon & Luck, 2014) and filtered offline using a band-pass filter of 0.1–30 Hz. EEG triggers were time-locked to the onset of the critical word and data was epoched from -200 ms prior to word onset to 800 ms post word onset using a baseline correction. Since our sentences were natural utterances of different durations, our EEG triggers were hand-defined according to onset of the critical word in the stimulus in milliseconds (average onsets for Standard Phraseology triggers were 0.64 ms, Nonstandard Phraseology 0.62 ms, and Incongruent Phraseology 0.62 ms).

Eye-blinks were removed by running an Independent Component Analysis (ICA) using the analysis implemented in EEGlab as well as the procedure described by Nunez and colleagues (2016). After eye-blink correction, trials were rejected according to the following criteria: we utilized an automatic artifact rejection procedure with a 200 ms sliding window and 50 ms steps. If any of the electrodes exceeded a ± 100 μV threshold within this time-window, the epoch was excluded from the analyses. Very noisy electrodes were interpolated by using the surrounding electrodes to reconstruct the signal.

ERP Data

The data, analysis code, and materials to reproduce all analyses and figures are available on OSF. https://osf.io/w489e/overview?view_only=731c46c435cd499fbda5b437dad2622b

The pre-processing pipeline led to the exclusion of 153 trials which comprised 10.62% of the data. A linear mixed-effects model[e] was fit to EEG voltage with Condition and Latency Window and their interaction as fixed effects, and Participant included as a random effect. Bonferroni-corrected pairwise comparisons between the three conditions revealed significant differences in multiple latency windows (see Figure 4 and Table 2). In the N400 window (300-400/400-500 ms), all pairwise comparisons were highly significant (p<.0001*), indicating robust differentiation between all conditions (see Figures 5 and 6). There was no effect of operational role (ATCO/Pilot) on EEG voltage (Est.=-0.26, SE=0.144, p=0.67).

---

[e] Voltage ~ Cond*Latency Window (1|Participant)

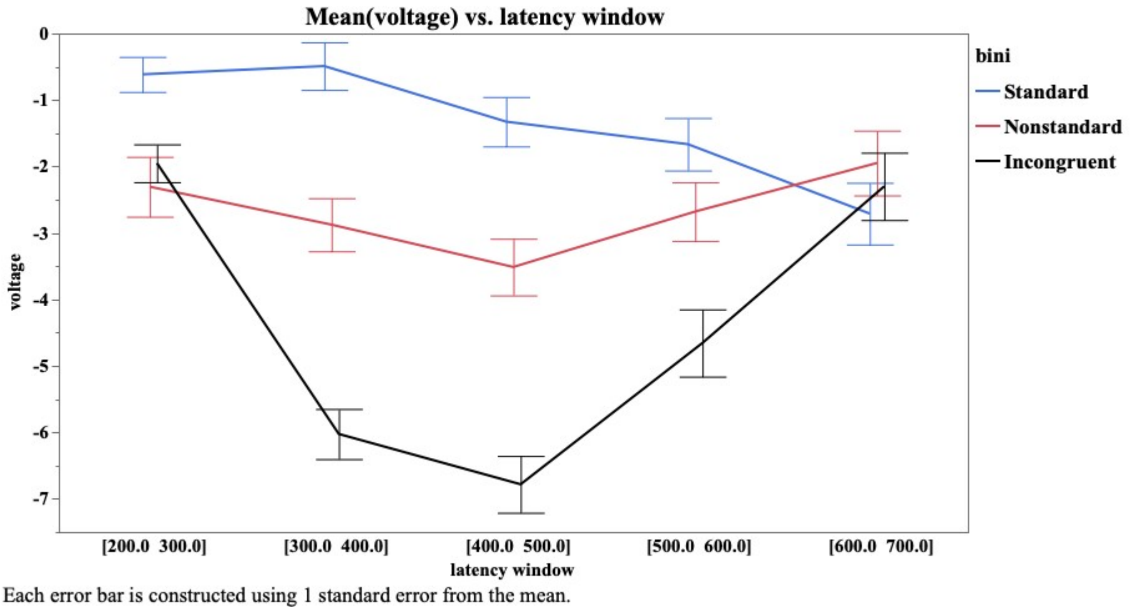

Figure 4:Mean EEG amplitude comparing the three experimental conditions in the parietal ROI, 200–700 ms post-stimulus onset.

Table 2: Pairwise comparisons (Bonferroni-corrected, *denotes p<.009, ** denotes p<.0001, n.s. denotes p>0.05).

Latency Window (ms)	Standard-Nonstandard	Standard-Incongruent	Nonstandard-Incongruent
200–300	*	*	n.s.
300–400	**	**	**
400–500	**	**	**
500–600	n.s.	**	**
600–700	n.s.	n.s.	n.s.

Figure 5: topographical maps for the three experimental conditions in the critical N400 time window: 300–500 ms post-stimulus onset.

Figure 6:Brainwaves time-locked to the critical word onset showing the three experimental conditions. Electrode is taken from the parietal ROI used in the analyses.

Behavioural Data

In addition to EEG data, we also collected response data for the recall task. Cleaning procedures resulted in the exclusion of 84 trials (6.08% of the data, N=1370). A linear mixed-effects model on reaction times was fit to reaction times with Match (Match vs. Mismatch), Condition (Standard vs. Nonstandard vs. Incongruent), and their interaction as fixed effects. The optimal model[f] contained log-transformed reaction times, an interaction between Match and Condition, and random intercepts for participant. There was a significant difference between this model and a model in which the interaction was removed, indicating the model containing the interaction was the best fit ($\chi^2(2)=14.62$, p=0.0006*).

Pairwise comparisons revealed that reaction times did not differ significantly between matched and mismatched trials in the Standard Phraseology condition (Est.=-0.05, SE=0.03, t= -1.7). However, they did differ significantly in the Nonstandard condition (Est.= -0.14, SE=0.03, t= -4.8*), where matched trials resulted in much faster reaction times than mismatched. Likewise, reaction times also differed significantly in the Incongruent condition (Est.= -0.21, SE=0.03, t= -7.2*), where again matched trials were faster than mismatched (see Figure 7).

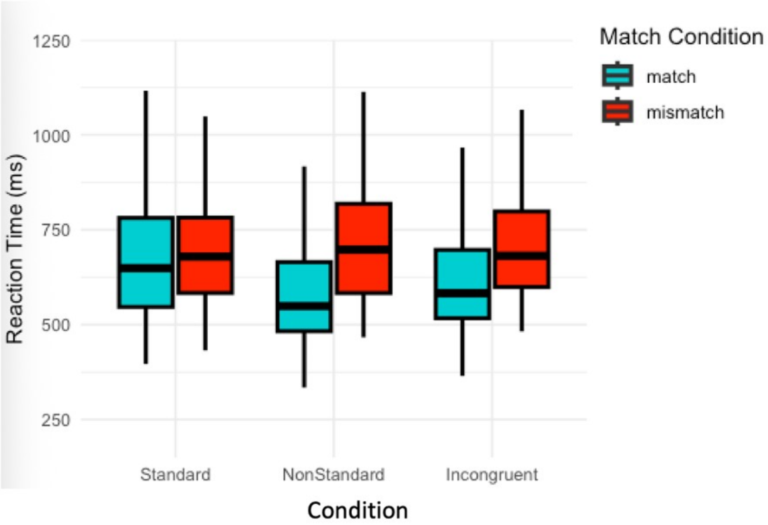

Figure 7: Reaction times (in ms) for each utterance type and match condition

---

[f] Logreaction_time ~Match*Cond+(1|Participant)

Accuracy rates were analysed as a function of Condition (Standard vs. Nonstandard vs. Incongruent) and Match (match vs. mismatch). A generalised linear model[9] (binomial distribution, link: logit, firth-adjusted maximum likelihood) revealed a weakly significant effect of Condition on accuracy ($\chi^2(2)=6.31$, $p=0.042$)* and a pairwise comparison showed that accuracy differed significantly between matched and mismatched trials in the Incongruent condition ($\chi^2(1)=25.04$, $p<0.001$*). This indicated that participants were substantially less accurate when the recall word mismatched the auditory stimulus (e.g. *Remain this jungle* -- *PAPER*). There were no significant differences between Match and Mismatch for Standard Phraseology (no errors occurred for these trials) or Nonstandard Phraseology ($\chi^2(1)=1.22$, $p=0.26$).

Discussion

Our results show that violations to monoreferentiality do indeed affect language processing in the expert listener. Aviation experts exhibited an N400 effect for both Nonstandard and Incongruent stimuli, with the Incongruent condition eliciting stronger negative amplitudes than the Nonstandard. This effect was strongest in the 300-400 and 400-500 ms latency windows (i.e. the N400 window), after which point the effect dissipated. This pattern suggests that experts rapidly evaluate incoming phraseology against their stored expectations for the domain-specific language of Aviation English. The graded response for Incongruent vs. Nonstandard indicates early sensitivity not only to outright semantic incongruity, but critically- to nonstandard aviation phraseology.

When listening to spoken language, we rely on a combination of context and predictability to aid us in successful sentence processing. The context of a sentence helps us to judge what word or phrase comes next, and predictability helps us to select the item that is most likely to fit the sentence. Thus, *a female chicken is a* _____ leads to the predictable closure of *hen* in

---

[9] glm = Correct ~ Condition*Match

most cases. That said, expertise can affect the predictability of this closure, and a poultry farmer or specialist may choose to complete the sentence with the word *pullet*, denoting a young female chicken. Expert language differs from general language in a myriad of ways, but the most prominent feature is a severely constrained sentence context. In Aviation English, this is translated through the monoreferential status of lexemes in the language: while our general lexicon allows for the exchange of semantically related items such as *maintain, stay at, remain, preserve, sustain*, the domain specific lexicon of Aviation English only allows *maintain.*

In this study, we hypothesised that neural responses known to be receptive to linguistic errors (i.e. the N400, P600) would also be receptive to errors in expert language, but that we would also see an interaction between the strength of these responses and the type of error (nonstandard vs. incongruent). Our expectation was that this interaction would reflect conflict between general language and domain-specific language; simply put, nonstandard items from the general lexicon would still be active during the processing of Aviation English.

Our results showed precisely that: the N400 effect for sentences containing a nonstandard word or phrase (e.g. *go down to flight level 150*) was weaker than that for sentences containing completely incongruent words or phrases (e.g. *run over to flight level 150*). If the effect had been similar in strength, then this would have indicated that the general lexicon was not active during the processing of domain-specific language (cf. Federmeier & Kutas, 1999). Thus, our results indicate that our paradigm is effective, and that the observed effect is robust. This is reassuring, since auditory EEG can be more noisy than visual EEG due to the temporal and acoustic variability of the auditory signal, environmental and physiological changes, and baseline/overlap issues (cf. Kutas & Van Petten, 1990; Hagoort & Brown, 2000; Pawlowski et al., 2019).

Paper Type: Full paper/article: 12-to-20-pages

This project was borne out of almost a decade of discussions and consultations between a neurolinguist and an aviation language practitioner. Early on in our conversations, we recognised that both of our fields were suffering from an absence of neural data on expert language processing. Our findings suggest that linguistic expertise involves not only the acquisition of new terminology but also the development of parallel, context-sensitive representational pathways for domain-specific words: simply put, **expertise fundamentally reshapes language processing.** This collaboration has been instrumental in the design and development of our current work, and we believe that the most effective way to ensure continued comprehensive research outputs is to hear from the people who work with the language in real, day-to-day situations.

The ultimate aim of these collaborations is to advance safety research by bringing rigorous scientific methodologies to the study of aviation communication. Not only do questions remain about the linguistic implications of these experimental findings, but we must also ask ourselves to what extent can they inform language pedagogy and assessment, particularly within specialised domains such as Aviation English? Currently, the field of Aviation English training and assessment lacks hard evidence on the impacts of deviations from standard phraseology on the brain, and the field of linguistics lacks sufficient knowledge about the fields in which expert speakers process specialised languages. Moving forward, we aim to create a dynamic forum for exchanging insights that bridge theory and practice, and stimulate ongoing research collaborations far beyond this project.

Paper Type: Full paper/article: 12-to-20-pages

Acknowledgements

This publication arises from research funded by the John Fell Oxford University Press Research Fund and a British Academy & the Leverhulme Trust small research grant awarded to Dr. Hilary Wynne.

Paper Type: Full paper/article: 12-to-20-pages

References

British Academy and The Leverhulme Trust Small Research Grant (2025). *Eliciting language-specific signatures for expertise in the human brain* (Grant No. SRG2425\251626). https://www.thebritishacademy.ac.uk/funding/schemes/ba-leverhulme-small-research-grants/past-awards/british-academyleverhulme-small-research-grants-awards-2024-25/

University of Oxford and Oxford University Press John Fell Fund (2025). *Eliciting language-specific signatures for expertise in the human brain.*

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), 483–495. https://doi.org/10.1037/0278-7393.10.3.483

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Bryce, T. G. K., & Blown, E. J. (2012). The nature of astronomy expertise. *Research in Science Education*, *42*(2), 225–246. https://doi.org/10.1080/09500693.2011.642325

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55-81. https://doi.org/10.1016/0010-0285(73)90004-2

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1982). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Friginal, E., Mathews, E., & Roberts, J. (2019). *English in global aviation: Context, Research, and Pedagogy*. Bloomsbury Publishing.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition*, *66*(2), 115–152. https://doi.org/10.1016/S0010-0277(98)00020-1

Gobet, F., & Simon, H. A. (2010). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *34*(2), 352–362. https://doi.org/10.1207/s15516709cog2404_4

Gotti, M. (2008). *Investigating Specialized Discourse*. Peter Lang.

de Groot, A. D. (1965). *Thought and choice in chess*. Mouton. (Original work published 1946).

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, *38*(11), 1518-1530.

Howard, J.W. (2008). Tower, am I cleared to land?: Problematic communication in aviation discourse. *Human Communication Research* 34 : 370-91. https://doi.org/10.1111/j.1468-2958.2008.00325.x

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358. https://doi.org/10.1037/h0043688

Kirklik, A., & Byrne, M. D. (2019). *Cognition and decision making: Human factors in engineering and design* (2nd ed.). CRC Press.

Krifka, M., S. Martens, and F. Schwarz. (2003). Group interaction in the cockpit: Some linguistic factors. *Group Interaction in High Risk Environments.* Germany: Gottlieb Dalmer. 75-102.

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. 10.1038/s41598-024-56090-6

Kutas, M., & Van Petten, C. (1990). Electrophysiological perspectives on comprehending written language. *New trends and advanced techniques in clinical neurophysiology*, *41*, 155-167.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163. https://doi.org/10.1038/307161a0

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

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. https://doi.org/10.3758/s13414-020-01979-3

McMillan, D. (1998). *"Say again?" Miscommunication in Air Traffic Control*. M.Ed. thesis, Queensland University of Technology.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1166–1174. https://doi.org/10.1037/0278-7393.15.6.1166

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Re-examining the role of Broca’s area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience, 5*(3), 263–281. https://doi.org/10.3758/CABN.5.3.263

Pawlowski, G. M., Ghosh-Hajra, S., Fickling, S. D., Liu, C. C., Song, X., Robinovitch, S., ... & D'Arcy, R. C. (2019). Brain vital signs: expanding from the auditory to visual modality. *Frontiers in Neuroscience*, *12*, 968.

Pierce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203.

Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. A. (2002). On the constraints of encapsulated knowledge: Clinical case representations by medical experts and novices. *Cognition and Instruction, 20*(1), 27–45. https://doi.org/10.1207/S1532690XCI2001_2

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5*(3), 205–221. https://doi.org/10.1007/BF01323044

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*(4), 394–403.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, *5*(1), 29-46. https://doi.org/10.1016/0010-0285(73)90024-8

Simpson, R. (2022). *Statistics for Linguistics with R: A Practical Introduction*. Routledge. https://doi.org/10.4324/9781003118234

Troyer, M., & Kutas, M. (2020). Harry Potter and the Chamber of What?: The impact of what individuals know on word processing during reading. *Language, Cognition and Neuroscience*, *35*(5), 641-657. https://doi.org/10.1080/23273798.2018.1503309

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247-263. https://doi.org/10.1016/j.jml.2011.05.002

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. https://doi.org/10.3390/brainsci14101029

Wynne, H.S.Z. (2025). The psychological reality of ESP processing: Insights from Aviation English. *Selected Papers from the 2025 APLX, ETA & TESPA International Conference on English Language Teaching and Learning*. English Teachers' Association- Republic of China. https://ora.ox.ac.uk/objects/uuid:7c1a0b4d-7ca1-42ed-bc9a-243373184700/files/sv692t8842

Wynne, H.S.Z. and Emery, H. (2026). A glimpse into the expert lexicon: Behavioural evidence from Aviation English. *Acta Psychologica.* 268, 107031.

Wynne, H.S.Z., Cheng, J., & Emery, H. Eliciting language-specific signatures for expertise in the human brain: An exploratory EEG study on Aviation English. [under review] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6512396