Neuropsychologia 135 (2019) 107230

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Left occipital and right frontal involvement in syntactic category prediction: MEG evidence from Standard Arabic

Suhail Matar[a,*], Liina Pylkkänen[a,b,c], Alec Marantz[a,b,c]

[a] Department of Psychology, New York University, USA
[b] Department of Linguistics, New York University, USA
[c] NYUAD Research Institute, New York University Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Language
Prediction
Syntactic entropy
Arabic
Visual M100
MEG

ABSTRACT

Though recent years have seen a growth in research on predictive processes in language comprehension, their scope and mechanisms remain partially elusive. While mechanisms involved in predicting specific words are relatively well understood, those underlying syntactic prediction are still unclear. In part, this is because of the difficulty in designing experiments that manipulate syntactic predictability while controlling other variables. In this MEG study, we achieved this with a manipulation of syntactic category predictability within fully well-formed expressions of Standard Arabic. Participants read sentences beginning with a subject-adjective context, in which the presence of at least one of two possible cues (gender-incongruity and/or an intervening relative pronoun) was sufficient for predicting a target word's syntactic category. Absence of both cues (i.e., congruent subject-adjective context with no relative pronoun) increased uncertainty about the target's syntactic category. Using source analysis, we compared activity evoked by targets with predictable and unpredictable categories in the occipital lobe. We found an interaction effect consistent with previous findings: in the primary visual cortex, an early evoked component (visual M100) is enhanced only when the syntactic category was unpredictable. We also compared responses to pre-target predictive and unpredictive contexts across five bilateral frontal and temporal regions. In the right-hemispheric frontal region, we found a temporal cluster (~230 ms after adjective onset), where unpredictive contexts elicited more activation than predictive contexts. By hypothesis elimination, we conclude that the most likely variable driving this effect is syntactic entropy. Our results show that predictive mechanisms recruited during reading also involve predicting upcoming syntactic categories, implicating at least two cortical regions: the left visual cortex and the right frontal cortex.

1. Introduction

Though natural and effortless for humans, processing language involves many mechanisms. Natural language expressions are complex and contain information on multiple levels. For example, in the sentence.

(1) The worker lost her key.

The word 'key' represents a particular concept or lexical item (an object used to lock and unlock doors), but it also represents a more abstract concept: it belongs to the syntactic category of nouns. Comprehending the meaning of (1) entails, among other things, accessing both facets of the word 'key'.

In recent years, a growing body of literature has provided evidence for the hypothesis of predictive processing during language comprehension – namely, that the brain generates predictions about upcoming content from contextual information, which then facilitate the processing of subsequent information (see Kuperberg and Jaeger, 2016, for a recent review). The accumulated evidence shows that predictive processes operate on several levels of linguistic representation. Behavioral (e.g., Warren, 1970 — 'phoneme restoration') and neurophysiological work (e.g., Gagnepain et al., 2012; Ettinger et al., 2014; Ylinen et al., 2016; see also Gwilliams and Marantz, 2015, for work on Standard Arabic) has provided evidence for predictive processing at the phonological level. There is also a considerable amount of evidence for prediction generation at the lexical level (e.g., Kutas and Hillyard, 1984; Lau et al., 2013; Lau et al., 2016; Maess et al., 2016b), showing that the

* Corresponding author. 10 Washington Place, New York, NY, 10003, USA.
E-mail address: matar@nyu.edu (S. Matar).

https://doi.org/10.1016/j.neuropsychologia.2019.107230
Received 1 March 2019; Received in revised form 25 September 2019; Accepted 11 October 2019
Available online 15 October 2019

degree to which a target word is predictable modulates brain activation levels as measured by electro- and magnetoencephalography (EEG and MEG, respectively), both before and after the onset of the target word.

At the syntactic level, which is the focus of this paper, several studies have also provided evidence for predictive processing – typically, for predicting the syntactic category (e.g., 'noun' or 'verb') of an upcoming target word. Initially, in an event-related potential (ERP) study, Friederici et al. (1993) showed that syntactic category violations (e.g., the preposition '*im*' in 'Der Freund wurde *im* besucht.', which translates as 'The friend was *in* the visited.') evoked an Early Left Anterior Negativity (ELAN) peaking at 180 ms after violation onset (See also Neville et al., 1991, for similar enhancement of the ERP component N125 as a result of phrase structure violations.). The ELAN was initially interpreted as evidence for syntactic structure building occurring at the time window corresponding to this component (Friederici, 2002).

However, in another EEG study, Lau et al. (2006) found that this early response is modulated by how predictable a word's syntactic category is; a sentential context that was not particularly predictive of a target word's category caused an attenuation of the ELAN on the target word, compared to a context that was highly predictive. This suggested that syntactic category expectations played a role in the ELAN phenomenon.

A string of ERP studies supports the hypothesis that such an early effect reflects a predictive process of a sensory nature (e.g., van Berkum et al., 2005; Dambacher et al., 2009; though see Nieuwland (2019) for a critical review). Subsequently, Dikker et al. (2009, 2010) presented evidence in favor of the *sensory hypothesis* as an explanation for the ELAN. In an MEG study, Dikker et al. (2009) showed participants sentences that were either grammatical, such as (2), or contained a syntactic category violation, as in (3).

(2) The boys heard Joe's stories *about* Africa
(3) The boys heard Joe's *about* stories Africa

Compared to (2), the violating word '*about*' in (3) caused a significant enhancement of an early occipital MEG-measured evoked component, occurring around 100 ms after word onset (Tarkiainen et al., 1999; we shall refer to this component as the visual M100). Moreover, the same effect was observed when comparing sentence pairs, in which the target word (such as '*reported*' in (4) and (5)) contained a strong visual cue (e.g., the '*-ed*' suffix) that identified the word's syntactic category, but not for bare stems that lacked these visual cues (such as '*tree*' in (6) and (7)):

(4) The discovery was *reported.*
(5) The discovery was in the *reported.*
(6) The owl was in the *tree.*
(7) The owl was *tree.*

In a subsequent MEG study, Dikker et al. (2010) showed that the visual M100 modulation was not exclusive to target words that contained a closed-class set of morphemes, which are strongly indicative of a syntactic category, but is also sensitive to form typicality – a measure of how indicative a word's *form* is of its syntactic class (Farmer et al., 2006). The same M100 effect was observed when the violating words were bimorphemic nouns with clear orthographic nominal cues (e.g., '*princess*') or form-typical monomorphemic nouns ('*soda*'), but not when they were form-neutral nouns ('*infant*'). Simply put, if a word's syntactic category is salient because of its form, the word drives a larger visual M100 when its syntactic category is unexpected.

The emerging picture from the above literature is one where activity in the visual cortex is enhanced by the appearance of a word from an unexpected (and violating) syntactic category, around 100 ms after word onset. This effect is found only when there are visual cues disclosing the word's syntactic category, and importantly when the preceding context is highly predictive of a syntactic category that is different from the one encountered.

However, there are two questions that this picture leaves unanswered. The first question is whether the early sensory effect exists outside the realm of violation paradigms, in the processing of grammatical sentences. All of the aforementioned M100 and ELAN studies relied on a violation paradigm: They all observed an effect when comparing grammatical sentences to ungrammatical stimuli. As such, it remains unclear whether the effect forms part of the processing of grammatically well-formed sentences, or is simply due to stimulus ungrammaticality. In order to adjudicate between these hypotheses, we need to test whether, in a fully grammatical paradigm, early activity in the visual cortex is sensitive to the predictability of a target word's syntactic category. But it is difficult to generate grammatical contexts that dissociate syntactic category predictability from other factors, especially semantic predictability. In this study, we use an adjectival modification paradigm in Standard Arabic, which allows us to vary syntactic category predictability without sacrificing grammaticality, and independently of semantic factors.

Provided that the early modulation in visual cortex activity is indeed observed in the processing of grammatical sentences, the second question concerns how these syntactic predictions are generated, and with what spatiotemporal profile.

There are two main variables from information theory that potentially explain the generation of a syntactic prediction effect: entropy (uncertainty) reduction and entropy (Hale, 2006, 2016). If entropy reduction is what drives the prediction, then a predictive context should evoke more activity than an unpredictive context, because the former involves more disambiguation (a greater entropy reduction). This has been dubbed the *entropy reduction hypothesis* (Hale, 2006; Linzen and Jaeger, 2015). On the other hand, if entropy is what drives the syntactic prediction, then an unpredictive context, which is more entropic, should evoke more activity than a predictive one. This is termed the *competition hypothesis*: entropy is associated with the level of competition between the different possible predictions (Linzen and Jaeger, 2015; Elman et al., 2005). Note that both variables have been shown to correlate with measures of neural activation (Entropy reduction: Nelson et al., 2017a. Entropy: Willems et al., 2016; Nelson et al., 2017b) and behavioral measures (Linzen and Jaeger, 2015) — even outside the direct realm of language studies (Schiffer et al., 2012, show evidence of entropy's influence on hippocampal blood oxygen-level dependent, or BOLD, signals during learning.)

The spatiotemporal profile of syntactic category predictions also remains unclear. Even if the visual cortex is sensitive to syntactic category predictability, it is widely considered to be a region where visual stimuli are processed in a bottom-up fashion. There is indeed evidence for preference to letter strings or sensitivity to string length in parts of the visual cortex (Tarkiainen et al., 1999), but it is highly unlikely that this same region generates these highly specific linguistic predictions.

Recent MEG studies have investigated neural correlates of generating lexical predictions. Dikker and Pylkkänen (2013) contrasted pictorial contexts that were more, or less, predictive of a specific upcoming noun. They found evidence for lexical pre-activation *before* the appearance of the noun, with the predictive context evoking more activity in the left mid-temporal cortex, the left ventro-medial prefrontal cortex, and the visual cortex. In another visual MEG study, Fruchter et al. (2015) examined the correlation between responses evoked by adjective-noun phrases, and the numeric transition probabilities between the two words (e.g., 'steel' follows 'stainless' with high probability, unlike 'stance' following 'tough'). They found that this metric explained activity in the left middle temporal gyrus. Additionally, Maess et al. (2016b) used auditory sentential contexts in which verbs were predictive or unpredictive of specific sentence-final nouns. They found stronger N400s on the predictive (vs. unpredictive) verbs and the unpredicted (vs. predicted) nouns, and a negative correlation between the two N400 effects, localized using MEG to the left superior and middle temporal cortices.

Furthermore, in a functional Magnetic Resonance Imaging (fMRI) study, Willems et al. (2016) tested whether brain areas are sensitive to different measures of lexical prediction. They had participants listen to natural stories, in which information-theory metrics were calculated for each word. Specifically, the surprisal and entropy of each word was calculated using a trigram model (i.e., considering the two preceding words). Surprisal is a function of the probability of encountering a word, given its preceding context, with words that are less likely to occur being more surprising. Lexical entropy, or uncertainty, is a function of the probability distribution of the upcoming word, based on the current context, with the current word less entropic if it is more predictive of the upcoming word. They compared natural stories to reversed speech segments, and found left- and right-frontal sensitivity to entropy, and bilateral temporal and right subcortical sensitivity to surprisal.

The fact that cortical areas are sensitive to lexical metrics such as transition probabilities (Fruchter et al., 2015), entropy, and surprisal (Willems et al., 2016) suggests that the brain should be able to generate predictions on a lexical level. If that is the case, and provided that sensory areas are indeed sensitive to syntactic category predictability, can we find evidence for cortical sensitivity to measures of syntactic information?

Another fMRI study by Henderson et al. (2016) used a paradigm similar to the one employed in Willems et al. (2016), but focused on syntactic surprisal in reading: each word in the texts used was parsed for how surprising its syntactic category is, based on the preceding context. The words were binned into high and low surprisal groups. They found increased BOLD signals for high-as opposed to low-surprisal words in the left inferior frontal gyrus (IFG) and the left anterior temporal lobe (ATL). Additionally, they identified a broader network, in which the BOLD signal correlated positively (bilaterally in the IFG and the fusiform gyrus) or negatively (right middle frontal gyrus) with syntactic surprisal.

Bonhage et al. (2015) recently conducted a fMRI study comparing real and pseudoword sentences that were predictive of either a verb or a noun in the final slot. The rationale was that real sentences would facilitate generation of both lexical predictions and syntactic category predictions, whereas the pseudoword sentences would only generate a prediction for the syntactic category. They reported that bilateral inferior frontal cortices showed higher BOLD signals for pseudoword sentences compared to real sentences, whereas bilateral temporal and insular regions showed higher BOLD signals for real sentences compared to pseudoword sentences. Note, however, that pseudowords are not necessarily devoid of semantic content, and they might still trigger some form of lexical access or activate phonological or orthographic neighborhoods. They might also involve processes that are not part of real sentence processing. Therefore, the comparison between real and pseudoword sentences is likely associated with more than simply the different levels of predictions.

More recently, Matchin et al. (2017) conducted a fMRI study contrasting natural and pseudoword stimuli that could be full sentences, consecutive two-word phrases, or jumbled word/pseudoword lists. Only the sentence conditions could involve syntactic predictions. They found that, compared to the two-word phrases and the word lists, sentences (both real and pseudoword) corresponded to an increase in BOLD signal in the left IFG and the left posterior superior temporal sulcus (pSTS). It is worth noting, however, that apart from their use of pseudowords, the contrasts in the experiment were between sentences and less natural conditions (consecutive two-word phrases and jumbled word lists). The question remains, then, as to whether the same result would be observed when contrasting strictly sentential contexts that are differentially predictive at the syntactic level.

These fMRI studies suggest a large network might be sensitive to syntactic information metrics. In our study, we used the regions highlighted in this literature to test whether the brain generates syntactic predictions and, if so, which metrics are likely associated with that process.

To sum up, to our knowledge, no study has thus far used a paradigm that isolates syntactic category predictability in a grammatical sentential context. An ideal paradigm to study this phenomenon would have: (i) grammatical sentential contexts; (ii) the ability to vary syntactic category predictability independently, and (iii) maximally matched pre-target content. This study uses a Standard Arabic paradigm that meets all these prerequisites.

1.1. *Adjectival modification in Standard Arabic*

In Standard Arabic, adjectival modification is typically post-nominal: the adjective occurs after the noun being described. Post-nominal adjectives, as in (8a) and (8b), usually agree with the described noun in terms of gender. (F and M indicate feminine and masculine, respectively. Additionally, following linguistic conventions, hyphens denote morphemic boundaries, whereas periods separate multiple meanings encompassed in a single morpheme). Generally, feminine gender in the singular is expressed by a single-letter suffix (compare: المفقود, the masculine version of the adjective 'missing', and المفقودة, which is the feminine version).

(8)	a.	al-ʻa:mil	l-mafqu:d	
		the-worker.M	the-missing.M	
		'the missing (male) worker'		
	b.	al-ʻa:mil-**a**	l-mafqu:d-**a**	
		the-worker-F	the-missing-F	
		'the missing (female) worker'		
	c.	al-ʻa:mil-**a**	l-mafqu:d-**a**	sayya:r-**a**-tuha
		the-worker-F	the-missing-F	car-F-hers
		'the (female) worker whose car is missing'		
	d.	al-ʻa:mil-**a**	l-mafqu:d	mufta:ḥ-uha
		the-worker-F	the-missing.M	key.M-hers
		'the (female) worker whose key is missing'		

However, Standard Arabic has another adjectival construction, which we will term the Complement Adjective Phrase (CAP). In this construction, which appears in (8c) and (8d), an adjective is sandwiched between two nouns, but it describes the noun *following* it (the car or the key); it cannot be attributed to the first noun, meaning that we cannot deduce in (8c) or (8d) that the worker is missing. Indeed, the adjective in a CAP construction must agree with the second noun in terms of gender, but not necessarily with the first noun — as in (8d).

This is the key to the manipulation we used in this study. When a comprehender perceives a sentence that begins with a gender-congruent subject and adjective, as in (8b), there is more than one possible continuation that they can expect, including: (i) a verb, (e.g., 'The missing worker criticized … '), or (ii) a second, gender-congruent noun in a CAP construction, as in (8c). However, if the adjective does not match the subject in terms of gender, as in (8d), then there is only one possible continuation following the adjective: Because of the gender incongruity between the subject and the adjective, what follows must be a second noun whose gender matches the adjective's, just like (8d); it is the only possible grammatical continuation. We can say that the adjective in (8c) is associated with more uncertainty or entropy regarding the upcoming syntactic category, compared to the adjective in (8d), which is highly predictive.

Importantly, the nouns 'car' and 'key', in (8c) and (8d) respectively, can be replaced by any nouns of their respective genders. Thus, if a prediction about the word following the adjective of a CAP construction is indeed generated, it is a more general prediction for the syntactic category of nouns, and not for a specific lexical item.

Interestingly, the CAP construction contains a relative clause ('the worker *whose* car is missing') but does not include an overt relative pronoun. However, if we replace the determiner attached to the adjectives in (8c) and (8d) with a standalone, overt relative pronoun, as in (9a) and (9b), the resulting phrases convey the exact same meaning, and have the same syntactic structure, since both sentences involve relative clauses. However, in both (9a) and (9b), the only possible word that can follow the adjective is a noun of a gender that matches the adjective's, just as in (8d). In addition, as before, 'car' and 'key' can be replaced with any nouns of their respective genders.

(9)	a.	al-ʿa:mil-a	llati:	mafqu:d-a	sayya:r-a-tuha
		the-worker-F	that	missing-F	car-F-hers
		'the (female) worker whose car is missing'			
	b.	al-ʿa:mil-**a**	llati:	mafqu:d	mufta:ḥ-uha
		the-worker-F	that	missing.M	key.M-hers
		'the (female) worker whose key is missing'			

So, while a gender-congruent adjective is unpredictive –as it can be followed by either a verb or a noun– a cue in the form of either gender-incongruity or an intervening relative pronoun (or both) generates a strong prediction for a noun in the Target slot.

Thus, what the examples above provide is an experimental design suitable for our questions, in which we can (i) manipulate the predictability of the upcoming syntactic category, (ii) while keeping lexical predictions constant, (iii) using a minimal manipulation, and (iv) without sacrificing grammaticality or resorting to violations.

1.2. The current study

In this MEG study, we used sentential contexts as shown in Fig. 1A. The sentences started with a Subject-Modifier pair and included two manipulations: (i) Gender-congruity/incongruity between the Subject and the Modifier/Target, and (ii) Type of Modifier: either adjective (CAP construction) or relative clause (relative pronoun + adjective). In three of these four conditions, the Modifier was predictive of the appearance of a noun in the Target slot; only when the Modifier was a gender-congruent adjective, was the Target's syntactic category unpredictable (Fig. 1D). In other words, the presence of at least one of two visual cues (gender-incongruity and/or a relative pronoun) resulted in a context that was highly predictive of a noun in the Target slot — these cues reduced the entropy in the Modifier window regarding the Target's syntactic category. However, the absence of both cues (i.e., in the congruent adjective condition) resulted in an unpredictive context, since the Target slot could be filled by either verbs or nouns — this Modifier is associated with a higher syntactic entropy. We used this design in a reading paradigm to ask two questions.

The first question, laid out in Fig. 1B, is whether we can find evidence in the visual cortex for syntactic category prediction that is not elicited by a violation. As discussed above, previous MEG results (Dikker et al., 2010) indicate that the visual M100 effect appears only if the words in question have a form which is typical of their syntactic category. In Standard Arabic, verbs and nouns correspond to distinct orthographic patterns (verbal patterns and nominal patterns; Saiegh-Haddad and Henkin-Roitfarb, 2014), which means that these patterns can be used as predictable visual cues that disclose a word's syntactic category. In other words, it should be theoretically possible for the parser to generate visual orthographic predictions regarding an upcoming syntactic category in Standard Arabic. If this is indeed the case, then in line with Dikker et al. (2009, 2010), we predicted an interaction between Congruity and Modifier Type. Since the Congruent Adjective condition is the only one in which a noun category in the Target is not predictable, the appearance of a noun in the Target slot should evoke an enhanced M100 response in the left visual cortex, compared to the other three conditions, in which a noun in the Target slot is highly predictable.

The second question we asked was contingent on the answer to the first question: If we do find evidence supporting early occipital sensitivity to syntactic category predictability, we can go back to the Modifier time window, and ask what the neural correlates of generating predictions about the Target's syntactic category are. However, the relative clause Modifiers are linearly and visually more complex, since they contain two standalone words, as opposed to one word in the adjective Modifiers. Because these visual differences are stark, they likely translate into differences in how the two types of Modifiers are processed — differences that are irrelevant to our purposes. Therefore, we mainly focused here on comparing the incongruent and congruent adjectival Modifiers alone (see Fig. 1C), though we also use the relative clause Modifier conditions as controls, as explained below.

We are ultimately interested in learning about how syntactic category predictions are generated, but in order to do so, we must take into account all the possible effects we may hypothetically encounter, even if they do not pertain to predictive processing. Though the congruent and incongruent adjectives differ by only a single letter, this difference could hypothetically elicit different effects related to bottom-up processing (Effects 1 and 2 in Fig. 1C), and to top-down predictive processing (Effects 3 and 4 in Fig. 1C).

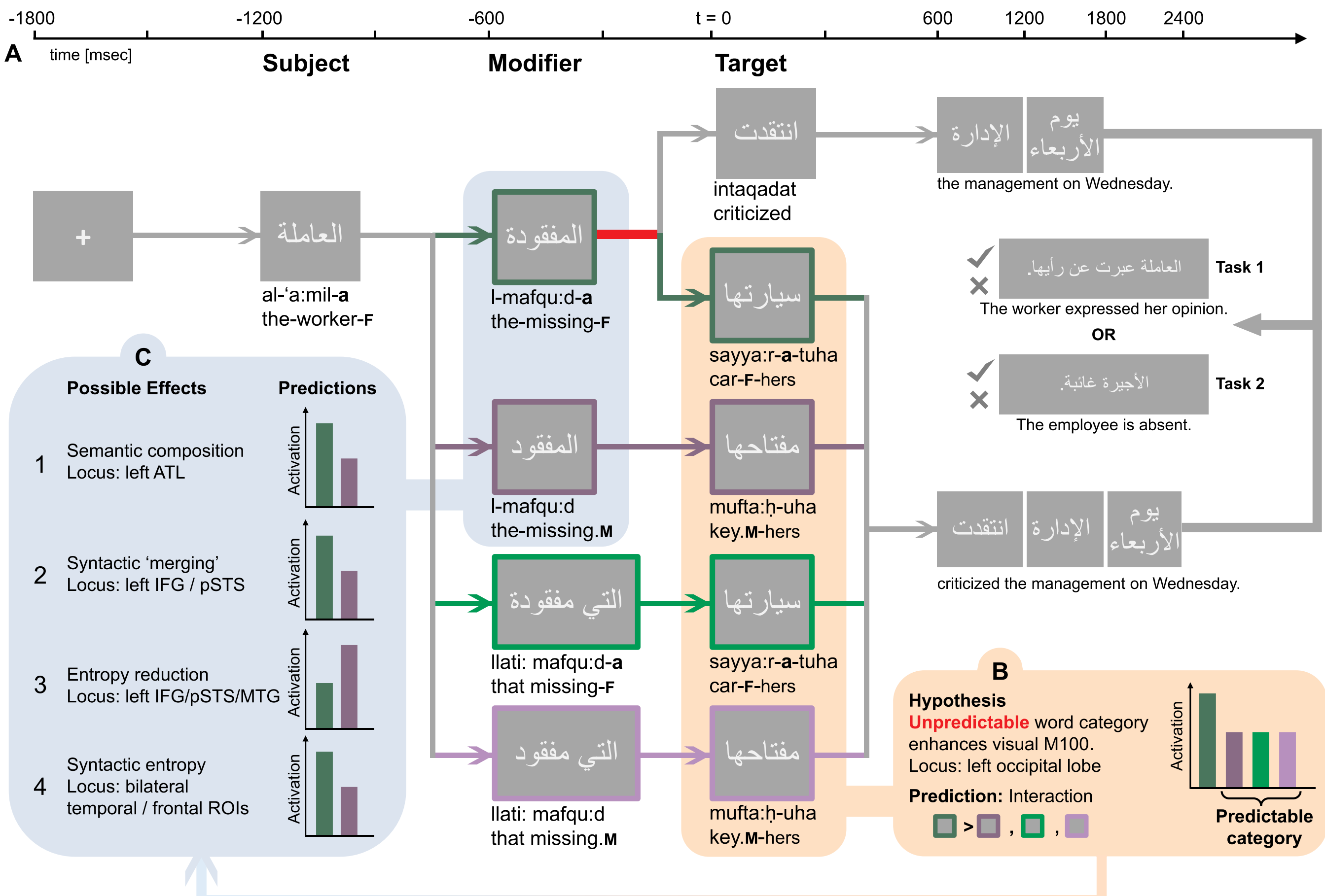
There is a possibility that when the Subject and Modifier are gender-congruent, they are readily composed semantically (Effect 1 in Fig. 1C) or 'merged' syntactically (Effect 2), unlike the gender-incongruent condition. If so, then based on the literature, we might expect a composition effect with more activity during the congruent condition in the left ATL at around 200 ms (as in Bemis and Pylkkänen, 2011; see also Westerlund et al., 2015, for results from Standard Arabic). For syntactic 'merging', we would expect more activation from the Congruent condition in the left IFG/left pSTS (Goucha and Friederici, 2015; Zaccarella and Friederici, 2015; see Zaccarella et al., 2017, for a meta-analysis. However, as one reviewer noted, ERP literature on feature agreement suggests that a gender agreement manipulation does not necessarily drive syntactic processing mechanisms—for a review, see Molinaro et al., 2011). Regarding the timing of a putative 'merge' effect, a recent MEG study by Flick and Pylkkänen (2018) found a left posterior temporal activation increase at around 200 ms for contexts that were more readily syntactically composable. Effects 1 and 2 do not involve any top-down syntactic category prediction. They simply reflect potential effects that we might encounter when investigating the Modifier time window. Unless we can rule out these semantic or syntactic composition interpretations as explanations for any effects found, we cannot deduce that the effects reflect syntactic category prediction.

Assuming the brain does engage in top-down syntactic category predictions, we can expect one of two possible patterns of activation, shown as Effects 3 and 4 in Fig. 1C. The first is that the predictive (incongruent) adjectives trigger pre-activation of the Target's syntactic category (as Dikker and Pylkkänen, 2013, have shown for lexical predictions), resulting in more activation in the left IFG and pSTS (Matchin et al., 2017). This would be in line with the entropy reduction hypothesis (Hale, 2006, 2016; Linzen and Jaeger, 2015). Importantly, any prediction generated in the incongruent adjective condition cannot be limited to the gender morpheme alone: gender morphology is different for nouns and verbs in Arabic. A nominal gender suffix must be bound to a nominal pattern. Therefore, generating a prediction about the gender suffix entails predicting the accompanying syntactic category.

The second predictive possibility is that the unpredictive (congruent) adjectives may evoke more activity than the congruent ones, driven by their elevated syntactic entropy. Because congruent subject-adjective contexts can be followed by Targets that are nouns or verbs, there is more uncertainty about the Target's syntactic category. However, the locus of such a putative effect is less clear, with the literature citing various bilateral frontal and temporal areas (Bonhage et al., 2015; Willems et al., 2016). If we find such a pattern in a spatiotemporal location that is not expected for any of the other effects, we can conclude that the entropy (or competition) hypothesis is the most likely explanation.

Lastly, the two relative clause Modifier conditions present an important control in this time window. The visual difference between congruent and incongruent conditions is the same across both types of Modifier: the single-letter gender morpheme. However, while in the adjective Modifier conditions, this visual difference generates a crucial gap in how predictive the context is of the Target word, that same difference is immaterial in the relative clause Modifier conditions. This is because the relative pronoun is a cue present in both congruent and incongruent conditions, which can be used to predict a Target noun. Thus, if any effects we find when comparing the two adjective Modifier conditions really do reflect a predictive process, then this effect should be eliminated when comparing the two relative clause Modifiers.

In summary, this MEG study uses a grammatical sentential paradigm in Standard Arabic that manipulates (i) gender-congruity between Subject and Modifier/Target, and (ii) Modifier Type, to investigate two questions. We expected an interaction between the factors in the visual cortex early on after Target onset, with more activation elicited by the congruent adjective condition, since its Target is the most surprising and unpredictable. If that is the case, we also expected to see differences in activation on the Modifier between the predictive and unpredictive contexts: barring any non-predictive interpretations, more activation for the predictive (gender-incongruent) context would lend support to the entropy reduction hypothesis, while more activation for the



D

Modifier	Gender Congruency	Condition	Example Gloss 'Translation'			Target category predictable?
			Subject	Modifier	Target	
Adjective	Yes	■ (dark green)	the-worker-F	the-missing-F	car-F-hers	No
			'The worker whose car is missing …'			
	No	■ (dark purple)	the-worker-F	the-missing.M	key.M-hers	Yes
			'The worker whose key is missing …'			
Relative Clause	Yes	■ (light green)	the-worker-F	that missing-F	car-F-hers	Yes
			'The worker whose car is missing …'			
	No	■ (light purple)	the-worker-F	that missing.M	key.M-hers	Yes
			'The worker whose key is missing …'			

Fig. 1. Experimental procedure and design. **A:** Trial structure. Words were presented visually and serially, each for 300 ms followed by a 300 ms blank screen. In all conditions of interest (colored) the appearance of a noun in the Target slot is predictable, except for the congruent adjective case (dark green), where there are at least two possible continuations (red line). Each trial was followed by one of two possible task types. F and M indicate feminine and masculine gender, respectively. **B:** Hypothesis corresponding to Target window. A Congruity x Modifier ANOVA is expected to reveal an early interaction in the left occipital cortex, with the syntactically uncertain condition (dark green) yielding an enhanced visual M100. If the prediction bears out, we can consider the Modifier time window. **C:** Possible effects we could find in the Modifier window. Several hypothetical effects are possible regarding activation differences between congruent and incongruent adjectives. **D:** Table showing the two-by-two design and manipulations, with example stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

unpredictive (gender-congruent) context would lend support to the entropy hypothesis.

2. Methods

2.1. Participants

We ran the experiment on New York University's New York (NY) and Abu Dhabi (AD) campuses. We recruited native Arabic speakers in both cities to participate in the experiment. Participants were right-handed with normal or corrected-to-normal vision. All participants provided written informed consent prior to data collection and were compensated for their time. In total, there were 39 participants. We excluded the data of 2 participants due to acquisition issues (one due to an extremely noisy recording, and another because most sensor channels were flat). In addition, we excluded the data of 10 other participants prior to analysis, due to low behavioral accuracy scores (<65%) on either of the two task types specified below. This left us with data from 27 participants (21 in New York and 6 in Abu Dhabi; 13 were female; ages: M = 30, SD = 9.03). These were from different parts of the Arab world: 13 Palestinians, 5 Lebanese, 3 Syrians, one Yemeni, one Emirati, one Jordanian, one Moroccan, one Kuwaiti, and one Egyptian.

2.2. Stimuli and task

We created thirty-six sets of stimuli corresponding to the design laid out in Fig. 1A. Four additional sets were created as practice sets, and were used before each experiment. Each set contained six trials: five trials followed the layout of Fig. 1A. A sixth trial, used as a filler to avoid the trials always having the same syntactic beginning, was as in (10):

(10)	ʻa:mil-at	l-maṣna'	intaqadat	l-'ida:ra	yawm l-'arba'a:'
	worker-F	the-factory	criticized	the-management	on Wednesday
	'The factory worker criticized the management on Wednesday.'				

Additionally, in each set, either the congruent or the incongruent relative clause condition featured a fronted noun, as in (11):

(11)	al-'a:mil-a	llati:	siyyar-a-tuha	mafqu:d-a	…
	the-worker-F	that	car-F-hers	missing-F	…
	'The worker whose car is missing … '				

These trials are equivalent in meaning to the non-fronted versions, but the shuffled order keeps participants from constantly expecting an adjective in the Modifier slot. Note that noun fronting is only possible for the relative clause conditions, not for the adjective ones. These filler trials were not included in our analyses.

In total, the experiment comprised 216 trials. Four additional sets were created for participants to practice with before data collection. In half the sets, the subject of the sentence was feminine, and in the other half, it was masculine. This determined the gender of the Modifier and Target nouns, depending on the condition. Importantly, this means that the distribution of the feminine morpheme marker was counterbalanced across congruent and incongruent conditions.

Moreover, since previous research has shown that the ERP's N400 is sensitive to animacy (Deutsch and Bentin, 2001), all subjects were chosen to be animate, whereas target nouns were inanimate. Additionally, we selected only adjectives that could be attributed to both the subject and the target nouns without changing the adjectives' meaning (e.g., 'missing' can be used as an attribute of 'worker', 'key', or 'car').

After each trial, a comprehension task phrase appeared. The participant's task was to indicate, by pressing one of two buttons, whether the task phrase is true or false based on the sentence presented during the trial. To ensure that participants were parsing the whole sentence, we included two kinds of task phrases, examples of which appear in Fig. 1A: (i) a task item which targeted the participants' understanding of the event in the sentence (e.g., 'The worker expressed her opinion.'), or (ii) a simple noun-adjective sentence, which targeted participants' understanding of the relation between Subject, Modifier, and Target (e.g., 'The employee is absent.'). The distribution of task types and correct answers was counterbalanced per condition, but the contents of the task items were not controlled. This is because the goal behind the task was simply to ensure the participants' engagement with the stimuli.

Table 1 shows the average frequencies of the adjectives and target nouns used in the experiment. For the target nouns, we calculated the frequencies with and without the possessive suffix. The data were obtained using python's *wordfreq* package (Speer et al., 2017). They are calculated using the Zipf scale (Van Heuven et al., 2014).

2.3. Experimental procedure

Prior to the experiment, each participant's head was digitized using a hand-held FastSCAN laser scanner (Polhemus, VT, USA). The digitized head was later used in the data preprocessing stage for co-registration purposes. Additionally, five points on the participant's head were marked with a marker then digitized: three on the forehead (center, left and right), and two anterior of each ear's auditory canal. Before entering the Magnetically Shielded Room (MSR), an explanation about the task was given to the participant, after which they completed a short practice session. Inside the MSR, marker coils were placed on the same digitized marker points, in order to localize each participant's head within the MEG. Marker measurements were obtained right before and right after the experiment, thus providing a measure of overall movement during the recording session.

The stimuli were divided into six blocks. Each block contained only one item from each set, for a total of 36 trials per block. The distribution of conditions from subsequent sets across blocks was done following a Latin square design, which ensured an equal number of all conditions per block. Then, the stimuli in each block were randomly shuffled. For each participant, the order in which the blocks were presented was also randomized.

We used a projector that relayed the image onto a screen inside the MSR; we made sure the visual angle across both NY and AD systems was the same, at approximately 0.7° vertically. We used PsychoPy2 (Peirce, 2007; version 1.84.2) and the Python Arabic Text Reshaper package (https://github.com/mpcabd/python-arabic-reshaper) for the presentation of the stimuli. Content was presented in white, against a gray background. Each trial began with a fixation cross that appeared in the center for 300 ms, followed by a blank screen for 300 ms. We used a rapid, serial, visual presentation (RSVP) mode: each word in a trial appeared in the center and had an on-screen time of 300 ms, followed by a blank screen of 300 ms, before the onset of the next word. To maximize similarity across conditions, the relative pronouns ('that') in the relative clause conditions were presented together with the following word, as in Fig. 1A. Another 300 ms separated the last word in a trial from the task

Table 1
Average frequencies (and standard deviation) of words used represented along the Zipf scale.

	Average Zipf frequency (standard deviation)	*t*-test statistic t(70) (*p*-value)
Congruent adjectives	3.73 (0.82)	−0.15 (0.88)
Incongruent adjectives	3.77 (0.82)	
Congruent target nouns	2.55 (1.18)	−0.2 (0.84)
Incongruent target nouns	2.61 (1.28)	
Congruent target nouns (no suffix)	4.21 (0.78)	−0.03 (0.98)
Incongruent target nouns (no suffix)	4.21 (0.92)	

sentence, which appeared all at once. After participants pressed either response button, the following trial began. Participants were instructed to focus during the experiment and to avoid blinking or moving as much as possible, especially during the RSVP.

During the experiment, MEG data were acquired continuously using a 157- and 208-channel (located in NY and AD, respectively) axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan), with a sampling rate of 1000 Hz and while applying an online low-pass filter of 200 Hz. The acquisition profile was maintained across both systems. The experimental runs lasted approximately 40 min, and participants were paid for their participation.

2.4. MEG data pre-processing

Prior to using the MEG results, we pre-processed the acquired data. First, the data were noise-reduced using the Continuously Adjusted Least Square Method (CALM; Adachi et al., 2001), provided in the accompanying MEG160 software (Yokohawa, Electric Corporation and Eagle Technology Corporation, Tokyo, Japan); this discounts noise recorded by three magnetometer reference channels, located away from the participant's head, from the data of interest. The noise-reduced data were imported into MNE-Python (Gramfort et al., 2014; version 0.14) and Eelbrain (Brodbeck, 2018; version 0.25.1), where they were band-pass-filtered between 1 Hz and 40 Hz and downsampled by 5. Following standard lab procedures, data from bad channels (flat or excessively noisy channels; number in NY: min = 7, max = 13, median = 8; number in AD: min = 8, max = 16, median = 13) were overridden and interpolated from the remaining sensors using MNE's implementation of the spherical spline interpolation method (Perrin et al., 1989).[1] After interpolating the bad channels, an independent component analysis (ICA) algorithm was then applied to the data. The ICA results helped identify and remove noise-related components based on visual inspection of the spatial and temporal profiles of these components. We only removed noise components that were identifiable (eye blinks and heartbeats), or characteristic of the MEG system. The data were then segmented into epochs, each spanning the whole sentence, from 100 ms before the onset of the first word right until the onset of the task item. Baseline correction was applied to each epoch based on the 100 ms of data that preceded each trial. Epochs containing signal amplitudes that exceeded a threshold of 3000 fT (NY) or 2000 fT (AD) were automatically rejected from the analysis. (The difference in thresholds is due to differences in ambient magnetic noise between the two systems and cities.) On average, 2.5% (SD = 3.6%) of the trials were rejected in this stage. Trials which participants answered incorrectly were also excluded, to avoid instances of incomplete or incorrect parsing of the sentences.

We scaled the FreeSurfer average brain (Fischl, 2012) to match each participant's digitized head shape, based on which we created a source-space consisting of 2526 vertices per hemisphere. Using the Boundary Element Model method, the activity at each vertex was used to calculate the forward solution. The inverse solution was then estimated for each subject, using an SNR value of 3. We opted for the unsigned, free orientation scheme, which imposes no constraints on dipoles' orientation with relation to the cortical surface. Estimates were obtained from the magnitude of each dipole, ignoring its orientation. The inverse solution resulted in a noise-normalized Dynamic Statistical Parameter Map (dSPM; Dale et al., 2000).

2.5. Regions of interest (ROIs) and test time windows

In line with previous literature (Dikker et al., 2009, 2010; Dikker and Pylkkänen, 2013), we expected to find, on the Target word, a modulation of the M100 component in the left visual cortex, with a higher peak for the unpredictable congruent adjective condition (Fig. 1B). Thus, we centered our analysis on a time window spanning 0–200 ms after target word onset, a window which housed the M100 component. We conducted the analysis in the left occipital lobe, dividing it into three Brodmann areas (BA): 17, 18, and 19 (see Fig. 2).

In the time window corresponding to the Modifier, we based our ROIs on the cortical regions reported in the literature surveyed in the introduction. This led us to carve five combined ROIs, located in the bilateral frontal and temporal lobes (see Fig. 3; the left temporal lobe was divided into an anterior and a posterior ROI, in order to be able to distinguish between Effects 1 and 2 in Fig. 1C). Frontal ROIs covered BAs 8, 9, 10, 44, 45, 46, and 47, whereas temporal ROIs included BAs 20, 21, 22, 37 and 38. It is worth noting that these ROIs, in addition to covering many of the reported cortical results in the literature, cover the space of possible processes that might be occurring during the Modifier time window (Fig. 1C). As for the timing of the effect, we expected composition effects to occur around 200 ms as explained in section 1.2. However, we did not have strong *a priori* expectations about the timing of effects related to the generation of syntactic predictions, since most of the relevant literature used fMRI with its poor temporal resolution. Thus, we opted for a wider test window encompassing −500 to −100 ms relative to the Target.

2.6. Data/statistical analyses

We analyzed the MEG data testing for contiguous clusters of timepoints, for which the statistics derived from the estimated activation levels were independently significant. We probed for these clusters in the activation timecourse obtained by averaging across all vertices within each ROI (referred to as a 'temporal test'). The cutoff for a timepoint contributing to a cluster was a *p*-value of .05. We set a minimum duration of 25 ms for cluster identification (unless otherwise indicated).

For all statistical tests reported, we conducted a subsequent cluster-based permutation test (following Maris and Oostenveld, 2007), comparing any resulting clusters against a distribution generated from the null hypothesis, based on 10,000 random permutations. For each permutation, each participant's condition labels were independently and randomly permuted. Additionally, since temporal tests included more than one ROI, we applied a False Discovery Rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

Finally, it is important to note that, following Sassenhagen and Draschkow (2019), we do not make any claims about the significance of the latencies of our clusters; we merely report the clusters identified, and any significant effects revealed by the subsequent cluster-based permutation tests within the test time window. The clusters' reported latencies and durations should be interpreted as having an approximate nature and should not be considered as claims regarding either the exact latency or duration of any effects.

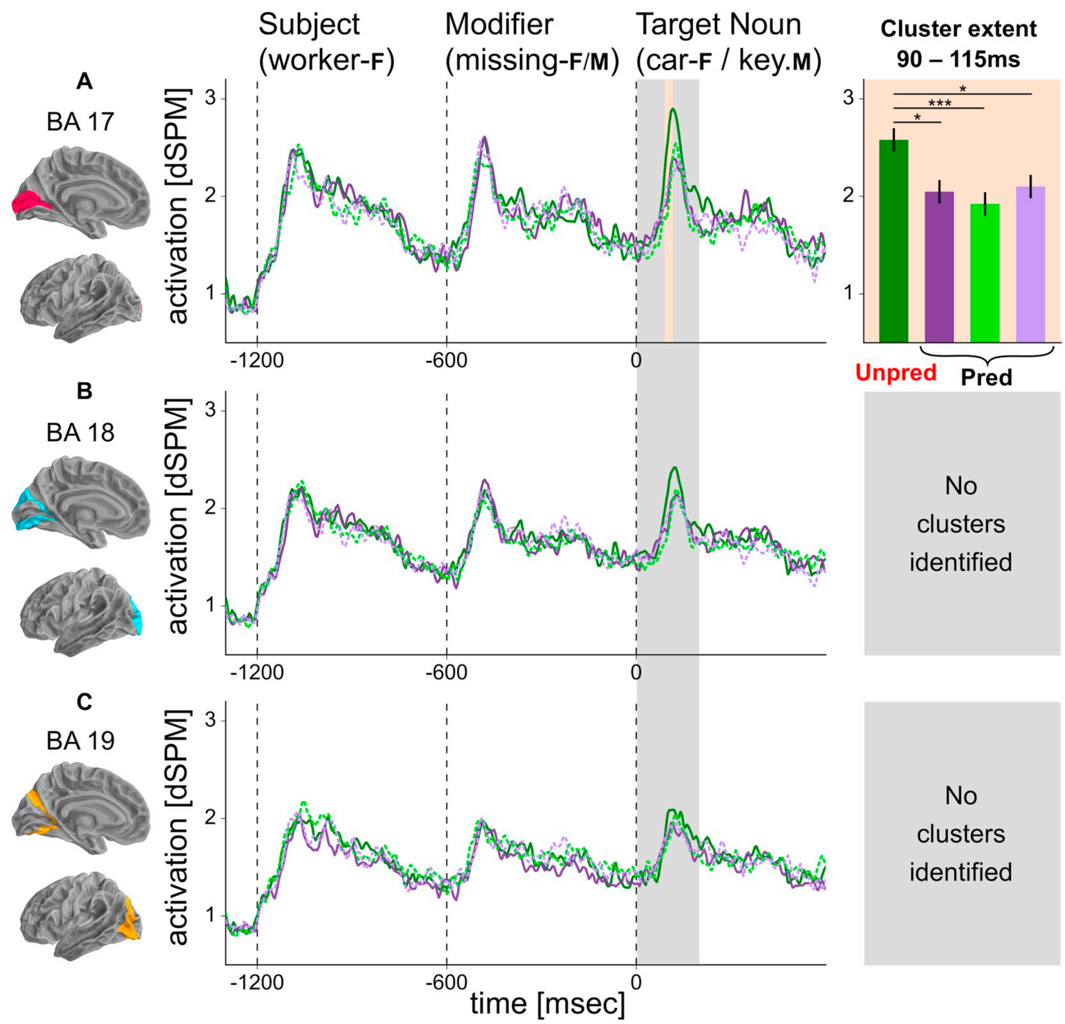
3. Results

3.1. Behavioral results

The 27 participants whose MEG data were included in the analyses had an average behavioral accuracy of 81.36% (SD = 6.23%).

Our behavioral data show that participants performed best on tasks following the incongruent relative clause condition (accuracy: 84.16%, SD = 7.58%), compared to congruent relative clause (74.90%, SD = 9.32%), the congruent adjective (78.91%, SD = 9.41%), and the incongruent adjective (76.75%, SD = 7.55%) conditions. However, these numbers cannot be used to compute statistics or be relied on to make any inferences about the difficulty of the different conditions per se, because the task items were not normalized for difficulty. We used

[1] One referee noted that, since the comparisons we are interested in are done in source space, it might have been better to remove, rather than interpolate, the bad channels.



Modifier	Gender Congruency	Condition	Example Gloss			Target category predictable?
			Subject	Modifier	Target	
Adjective	Yes		the-worker-F	the-missing-F	car-F	No
	No		the-worker-F	the-missing.M	key.M	Yes
Relative Clause	Yes		the-worker-F	that missing-F	car-F	Yes
	No		the-worker-F	that missing.M	key.M	Yes

Fig. 2. Analysis of occipital activation during Target presentation. Timecourses of estimated average activity in: **A:** BA 17, **B:** BA 18, and **C:** BA 19 in the left occipital cortex. t = 0 corresponds to the onset of the Target nouns. Shaded band (0–200 ms) indicates time window for the statistical 2 × 2 ANOVA and permutation test. Yellow band indicates a cluster. Bar plot (top right) shows activity levels per condition, averaged over participants and the time window corresponding to the identified cluster. Bottom panel is a legend, showing the Modifier type by gender congruency 2 × 2 design. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the task items simply to keep the participants engaged and to make sure they are processing the entire sentence, including the critical words. With that in mind, we did not analyze the behavioral results further.

3.2. MEG results: syntactic category predictability in the occipital lobe

In order to test whether evidence for syntactic category predictability could be found in the occipital lobe, we first ran a 2 (Modifier type: adjective, relative clause) by 2 (Gender-congruity: congruent, incongruent) within-subjects temporal ANOVA. We conducted our analysis in three subregions of the occipital lobe: BAs 17, 18, and 19, and on the time window corresponding to 0–200 ms after onset of the target word, which included the M100 component. The results appear in Fig. 2.

The test revealed a cluster in BA 17, or the primary visual cortex, extending from approximately 90 ms–115 ms after Target onset, and corresponding to the visual M100 peak (Fig. 2A). As expected, the

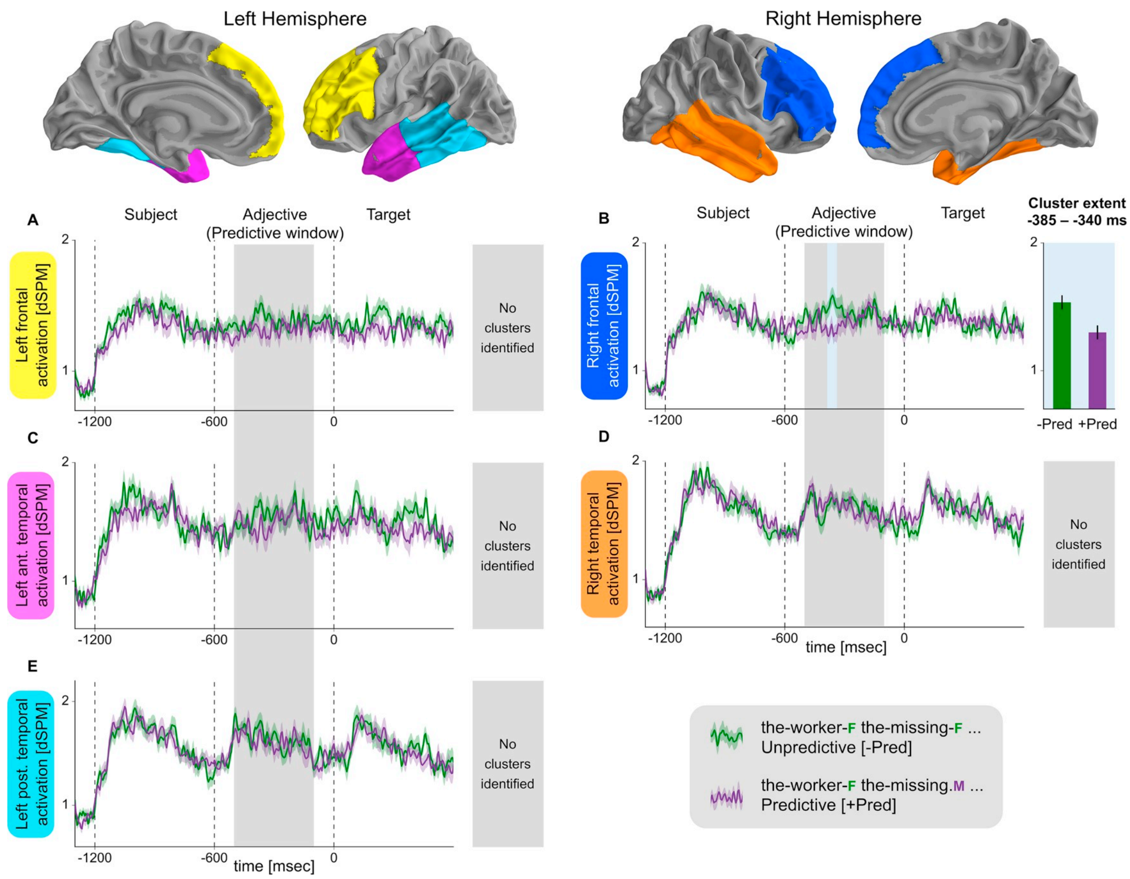


Fig. 3. Analysis of frontal and temporal activation in the adjectival conditions. Timecourses of estimated average activity in: **A:** left frontal; **B:** right frontal; **C:** left anterior temporal; **D:** right temporal, and **E:** left posterior temporal ROIs. Top panel: visualization of the ROIs used in each hemisphere. t = 0 corresponds to the onset of the Target. Gray bands (−500 to −100 ms) indicate test time window. Blue band indicates an identified cluster. Bar plot (**B**) shows activation levels per condition, averaged over participants and the temporal extent of the cluster. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

activation pattern within the cluster showed an interaction: the noun in the congruent adjective condition, which was least predictive of the Target's syntactic category, elicited the most pronounced Target M100, compared to the other three conditions, in which the appearance of a noun in the Target slot was equally highly predictable. The cluster-based permutation test indicated a significant effect in BA 17 within the test time window ($p = .036$, after correction for multiple comparisons).

3.3. MEG results: pre-target (modifier) time window

In addition to replicating syntactic predictability effects in the visual cortex, another question we were interested in was the generation of these predictions. Thus, we conducted another analysis on the time window preceding the Target.

For this analysis, we mainly focused on the trials in which the Modifier was an adjective (Fig. 1C), using a dependent *t*-test comparing congruent to incongruent adjectival Modifiers, in a time window corresponding to −500 to −100 ms relative to the Target. However, as a control, we also conducted the same analysis to compare the congruent and incongruent relative clause conditions, for which we did not expect to see any predictive processing effects. The temporal tests were conducted in the five ROIs seen in Fig. 3: left frontal, right frontal, left anterior temporal, left posterior temporal, and right temporal.

The test comparing the two adjectival conditions identified a cluster in the right frontal ROI (Fig. 3B), extending approximately from −385 ms to −340 ms, relative to Target onset. Within the cluster indicated, the average activation level for the congruent condition was higher than the incongruent condition. The cluster-based permutation test indicated a significant effect in the right frontal ROI within our test time window ($p = .0405$, after correction for multiple comparisons).

Though no clusters were identified in the other ROIs, we did notice some visual separation between the waveforms, especially in the left frontal and anterior temporal regions (Fig. 3A and C, respectively). To explore those separations, and to make sure no effects were overlooked due to the cutoff of 25 ms cluster duration, we conducted the same test again, this time allowing clusters as short as 10 ms. Other than the previously reported cluster, which did not change, the lower cutoff revealed additional clusters: one in the left frontal ROI (approximately −370 to −355 ms relative to Target), another in the left anterior temporal ROI (approximately −395 to −380 ms relative to Target), and two additional ones in the right frontal ROI (approximately −500 to −485 ms, and −415 to −395 ms after onset). However, the cluster-based permutation test failed to find any effects in any but the right frontal ROI ($p > .3$, after correction for multiple comparisons).

Additionally, we ran the same test, using the same ROIs and test window, but this time comparing the two relative clause Modifier conditions (Fig. S1). As expected, the test did not identify any clusters in any of the ROIs for these two control conditions.

Lastly, as an exploratory step, we conducted a whole-brain analysis using the same test parameters. Different clusters were identified, but the cluster-based permutation test did not reveal any significant effects ($p > .69$).

4. Discussion

In this MEG study, our aim was to investigate syntactic category predictions in grammatical sentences, using a reading paradigm. For that end, we used sentential stimuli in Standard Arabic, in a factorial design that varied syntactic category predictability using two manipulations: a gender-congruity manipulation (Subject and Target noun were either gender-congruent or incongruent), and Modifier Type (the Modifier of the subject was either an adjective or a relative clause). Though the manipulations involved minimal orthographic differences, these produced contexts that were differentially predictive of the Target's syntactic category. Presence of at least one of two possible visual cues (gender-congruity or a relative pronoun) was enough to create a context that fully predicted a noun in the Target slot. We wanted to investigate occipital sensitivity to this difference in category predictability and, if it exists, examine the generation of syntactic category predictions.

4.1. Early occipital sensitivity as evidence for syntactic category prediction

Our results from the analysis of the time window corresponding to the Target are consistent with previous findings that show early activity in the visual cortex – namely, a modulation of the visual M100 peak (Dikker et al., 2009, 2010, Fig. 2). In our experiment, the modulation is an interaction between the two factors in BA 17, such that when the Target's category is unpredictable, it elicits a more pronounced M100 peak, compared to predictable syntactic categories. The only significant differences in average levels of activations, as revealed by pairwise t-tests (bar plot in Fig. 2A), were between the congruent adjective condition and each of the other three conditions. It is worth examining two contrasts in this pattern.

The first contrast is between the congruent and incongruent adjective conditions. The pre-Target contexts of these two conditions are almost identical, the only difference being the gender morpheme. But note that the effect cannot be explained in terms of this morpheme alone, since its distribution was counterbalanced across the sets in the design. However, there may be an alternative explanation for this finding, which does not involve a purely syntactic prediction. It is possible that the context preceding the Target (Subject and Modifier) still generates predictions for a set of possible upcoming lexical items, beyond any general syntactic category predictions.

But if we examine the contrast between the congruent adjective and congruent relative clause conditions, which the pairwise t-tests revealed to be the most significant in the interaction pattern (bar plot in Fig. 2A), the abovementioned alternative explanation does not hold. These two conditions feature the same Target nouns (Fig. 1B), are both congruent with respect to the subject of the sentence, and are both preceded by the same set of adjectives. Moreover, the two conditions have equivalent meanings and correspond to syntactic structures of the same complexity. The only difference between the two (determiner vs. relative pronoun) appears before the Target is presented. Therefore, even if these contexts generate accompanying lexical predictions, their probability distributions would be identical. Thus, the visual M100 effect is best explained by syntactic category predictability. Aside from the previous literature, this is supported by the timing and location of the cluster.

It is also interesting to briefly consider the hitherto overlooked condition in which the Target is a verb (Fig. 1A). Following a congruent subject-adjective pair in Arabic, a noun is less commonly found (more unpredictable) than a verb. Indeed, the M100 induced by the Target verb is smaller than the M100 of the Target nouns. This supports the idea that the M100 is modulated by the unpredictability of the syntactic category. However, this pattern could also be due to different visual properties of nouns and verbs.

This result is consistent with a growing body of MEG (Dikker et al., 2009, 2010; Dikker and Pylkkänen, 2013) and EEG (van Berkum et al., 2005; Lau et al., 2006; Dambacher et al., 2009) literature that has provided evidence for early cortical sensitivities to lexical or syntactic predictability. Specifically, note that Dikker et al. (2009, 2010) showed that when a word's syntactic category is both unexpected (due to a violation of a syntactic prediction) and salient (because of its form), the visual M100 is larger. In this study, we have also shown a similar larger visual M100 for words whose syntactic category is both unexpected (here: due to an unpredictive context) and salient (because of the nominal patterns of Arabic).

However, it is important to note that in a recent review, Nieuwland (2019) revisited this literature and highlighted important drawbacks that led them to conclude that the evidence for the sensory hypothesis remains weak, pending replications. Central to their critique, Nieuwland (2019) invoked Tanner et al. (2015)'s important warning regarding using high-pass filtering at 1 Hz or similar in ERP studies, which runs the risk of smearing some components backwards in time. Unfortunately, we could not do away with high-pass filtering at 1 Hz, since our NY data are simply too noisy below this frequency. In response to Tanner et al. (2015), Maess et al. (2016a) write that "the application of a high-pass filter is reasonable if it improves the … (SNR) of the signal of interest, " and that the choice of cutoff frequency should be data dependent. We find that a cutoff of 1 Hz effectively rids us of low-frequency noise in our data. Nevertheless, high-pass filtering is problematic for many ERP components. For example, Tanner et al. (2015) simulate a P600 ERP component using an 800 ms-long cosine wave, corresponding to 1.25 Hz, and show severe distortion when high-pass filtering with a cutoff of 1 or 2 Hz. In our study, the word rate is 1.66 Hz (or 600 ms/word) and the components of interest have higher frequencies. In fact, the duration of the visual M100 deflections is approximately 100 ms, corresponding to about 10 Hz—well above our 1 Hz cutoff and outside the 2 Hz transition band.

More importantly, Tanner et al. (2015) show how an N400 can be smeared back in time if high-pass filtering with high cutoff values is used, creating an induced artefactual P2 and masquerading as an earlier effect. If we assume this is also the case in our study, and that the M100 modulation is simply an artefact of a backward-distorted N400-like component, we would expect to see a similar modulation on the Modifier window. This is because, in that window, our congruency manipulation should also produce an N400-like component that is distorted backwards in time. However, we find no evidence of this when examining the M100 of the Modifier time window (Fig. 2). Moreover, as one referee pointed out, our Target M100 deflection is the biggest in the time window of interest, making it further unlikely to be the result of an artefact caused by filter smearing. Therefore, for our purposes, it does not seem likely that the high-pass filtering caused any considerable temporal distortion of the pattern of activation.

To sum up, the effect reported here replicates previous MEG findings and provides evidence that the modulation of the M100 peak by word form predictability is part of the processing of well-formed sentences during reading. Additionally, because of our Arabic-specific design, we explain this effect in terms of occipital sensitivity to the predictability of the syntactic category. These sensory predictions should be especially viable in Standard Arabic, since verbs and nouns correspond to distinct orthographic patterns.

4.2. Neural correlates of prediction generation

The second aim of the study was to identify the neural correlates of generation of syntactic category predictions. To answer this question, we compared the activation patterns elicited by congruent (unpredictive/more entropic) and incongruent (predictive/less entropic) adjectives, across five ROIs corresponding to left and right frontal and temporal areas (Fig. 3).

Our analysis identified a cluster around 230 ms after Modifier onset in the right frontal ROI (Fig. 3B), with the more syntactically entropic condition eliciting more activation. Which of the possibilities laid out in the introduction (Fig. 1C) best explains this result, given the location and timing of the cluster?

The first possibility is that the congruent condition facilitates semantic composition between Subject and Adjective (Bemis and

Pylkkänen, 2011; Pylkkänen et al., 2014; see Westerlund et al., 2015, for results from Standard Arabic) compared to the incongruent one. If so, we expected an increase in activation for the congruent condition in the left ATL at around 200 ms after onset. Indeed, we did see a slight waveform separation at around that time in the left anterior temporal ROI, associated with a very small cluster, but the cluster-based permutation test failed to reject the null hypothesis in that area within our test time window.

A second possibility is that in the congruent condition, the Subject and the Modifier are readily 'merged' into a bigger syntactic unit, and that they are not 'merged' in the incongruent condition because of the gender mismatch. Previous literature has shown that the most likely locations to expect such an effect are the left IFG and the left pSTS (Zaccarella and Friederici, 2015; Pattamadilok et al., 2016; Zaccarella et al., 2017; Flick and Pylkkänen, 2018). A very small cluster was indeed found in the left frontal ROIs, but subsequent permutation tests failed to reject the null hypothesis.

A third possibility is that the incongruent condition, being more predictive of the Target's syntactic category than the congruent one, would reduce syntactic entropy, pre-activating the representation of that syntactic category. If so, we expected an increase for the incongruent activation in the left IFG and pSTS (Matchin et al., 2017), or even in the left middle temporal cortex (if it resembles lexical pre-activation: Dikker and Pylkkänen, 2013; Fruchter et al., 2015; Maess et al., 2016b). However, as discussed in section 3.3, though small clusters were identified in the left frontal or temporal ROIs, the permutation tests failed to reject the null hypotheses in these ROIs. Additionally, the entropy reduction hypothesis predicts a pattern that is the opposite of the one in the right frontal cluster we found. Therefore, it cannot explain our effect.

The remaining hypothesis is that the congruent adjective, because it carries more entropy regarding the Target's syntactic category, elicits more activity than the incongruent adjective. The cluster identified and the pattern observed in the right frontal ROI would be best explained by this hypothesis. Indeed, Bonhage et al. (2015) found that right frontal ROIs exhibited more BOLD signal for generation of purely syntactic (compared to combined syntactic and semantic) predictions.

Interestingly, Henderson et al. (2016) found that BOLD activity in the right middle frontal gyrus was negatively correlated with syntactic surprisal, which was the only information-theoretic metric considered in that study. The authors speculate that this negative correlation can be explained by an attention shift-based account, due to the reading paradigm the used, which involved eye movements. However, in our experiment: (i) our stimuli were presented word by word, minimizing eye movements, and (ii) we manipulated the information-theoretic metrics of entropy reduction and entropy, and found a positive correlation between activation in the right frontal ROI and entropy. We speculate, therefore, that what could have been underlying the negative correlation with surprisal observed in Henderson et al. (2016), was rather a positive correlation with another information-theoretic metric, such as entropy or a related metric.

It is also interesting to note the timing of the cluster, at approximately 230 ms after the onset of the adjective. In order for the entropy account to be valid, the brain should have access to sufficient information from the current word being processed, so as to generate predictions about upcoming material. According to the full decomposition model (Taft, 2004; Fruchter and Marantz, 2015), visual processing of a morphologically complex word involves decomposing it into individual morphemes, with evidence suggesting this happens around 170 ms after word onset (Solomyak and Marantz, 2010). Since the cues in our manipulations (the gender morpheme or the relative pronoun) are ultimately morphological in nature, this model suggests that the information required in order to generate a syntactic category prediction should be available approximately 170 ms after Modifier onset. This account fits well with the approximate latency of our cluster.

Finally, when the same comparison is made between the congruent and incongruent relative clause conditions (both of which are predictive of a noun in the Target slot), no clusters were identified in any of the regions (Fig. S1). However, the difference between these two conditions is identical to the difference between the congruent and incongruent adjective conditions — namely, the gender morpheme. This supports the idea that the effect we found in the right frontal ROI when comparing the adjective conditions is not simply due to the gender congruency per se, or the morphological difference, but rather due to what this difference entails in terms of predictive processing.

To sum up, the location, the approximate timing, and the pattern of activation in the cluster support an account of right frontal sensitivity to syntactic entropy and the competition hypothesis, according to which higher entropy drives more activation due to competing possible continuations. It might be the case that this information is relayed to other brain areas, eventually modulating the early visual responses to the next word.

5. Conclusion

In this MEG study, we examined predictive processing of syntactic categories during reading, using a fully grammatical design in Standard Arabic. We have extended previous findings by showing that early occipital activity –concretely, the M100 component in the primary visual cortex– is sensitive to syntactic category predictability; we have shown that this effect is also found for grammatical language and is not merely an epiphenomenon of previously used violation paradigms. Specifically, the M100 is enhanced when a word's syntactic category is less predictable. Additionally, we explored the neural correlates of the generation of these syntactic predictions in the pre-target time window. We found evidence for right-hemispheric frontal involvement at around 230 ms after onset of the predictive context, with unpredictive contexts eliciting more activity compared to predictive contexts. We considered a set of possible explanations for this effect and concluded that the most likely candidate is that it is driven by syntactic entropy.

Declarations of interest

None.

Acknowledgments

This work was supported by the NYU Abu Dhabi Research Institute (UAE) under Grant G1001 (AM). We would like to thank Julien Dirani and Samantha Wray for help with data collection in Abu Dhabi, Christian Brodbeck for Eelbrain clarifications, and Graham Flick, Maxime Tulling, and Yohei Oseki for helpful discussions. Additionally, we would like to thank the editor and two anonymous reviewers for providing valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2019.107230.

References

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., Ochiai, M., 2001. Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE Trans. Appl. Supercond. 11 (1), 669–672. https://doi.org/10.1109/77.919433.

Bemis, D.K., Pylkkänen, L., 2011. Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. J. Neurosci. 31 (8), 2801–2814. https://doi.org/10.1523/JNEUROSCI.5003-10.2011.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57 (1), 289–300. http://www.jstor.org/stable/2346101.

Bonhage, C.E., Mueller, J.L., Friederici, A.D., Fiebach, C.J., 2015. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. Cortex 68, 33–47. https://doi.org/10.1016/j.cortex.2015.04.011.

Brodbeck, C., 2018. Eelbrain 0.27 (Zenodo). Available at: http://eelbrain.readthedocs.io.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26 (1), 55–67. https://doi.org/10.1016/S0896-6273(00)81138-1.
Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., Jacobs, A.M., 2009. Event-related potentials reveal rapid verification of predicted visual input. PLoS One 4 (3), e5047. https://doi.org/10.1371/journal.pone.0005047.
Deutsch, A., Bentin, S., 2001. Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. J. Mem. Lang. 45 (2), 200–224. https://doi.org/10.1006/jmla.2000.2768.
Dikker, S., Rabagliati, H., Pylkkänen, L., 2009. Sensitivity to syntax in visual cortex. Cognition 110 (3), 293–321. https://doi.org/10.1016/j.cognition.2008.09.008.
Dikker, S., Rabagliati, H., Farmer, T.A., Pylkkänen, L., 2010. Early occipital sensitivity to syntactic category is based on form typicality. Psychol. Sci. 21 (5), 629–634. https://doi.org/10.1177/0956797610367751.
Dikker, S., Pylkkänen, L., 2013. Predicting language: MEG evidence for lexical preactivation. Brain Lang. 127 (1), 55–64. https://doi.org/10.1016/j.bandl.2012.08.004.
Elman, J.L., Hare, M., McRae, K., 2005. Cues, constraints, and competition in sentence processing. In: Tomasello, M., Slobin, D. (Eds.), Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates. Erlbaum, Mahwah, NJ, pp. 111–138.
Ettinger, A., Linzen, T., Marantz, A., 2014. The role of morphology in phoneme prediction: evidence from MEG. Brain Lang. 129, 14–23. http://10.1016/j.bandl.2013.11.004.
Farmer, T.A., Christiansen, M.H., Monaghan, P., 2006. Phonological typicality influences on-line sentence comprehension. Proc. Natl. Acad. Sci. 103 (32), 12203–12208. https://doi.org/10.1073/pnas.0602173103.
Fischl, B., 2012. FreeSurfer. Neuroimage 62 (2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.
Flick, G., Pylkkänen, L., 2018. Isolating Syntax in Natural Language: MEG Evidence for an Early Contribution of Left Posterior Temporal Cortex. BioRxiv, 439158 [Preprint]. Available from:
Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cogn. Brain Res. 1 (3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2.
Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6 (2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8.
Fruchter, J., Linzen, T., Westerlund, M., Marantz, A., 2015. Lexical preactivation in basic linguistic phrases. J. Cogn. Neurosci. https://doi.org/10.1162/jocn_a_00822.
Fruchter, J., Marantz, A., 2015. Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. Brain Lang. 143, 81–96. https://doi.org/10.1016/j.bandl.2015.03.001.
Gagnepain, P., Henson, R.N., Davis, M.H., 2012. Temporal predictive codes for spoken words in auditory cortex. Curr. Biol. 22 (7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015.
Goucha, T., Friederici, A.D., 2015. The language skeleton after dissecting meaning: a functional segregation within Broca's Area. Neuroimage 114, 294–302. https://doi.org/10.1016/j.neuroimage.2015.04.011.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Prakkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027.
Gwilliams, L., Marantz, A., 2015. Non-linear processing of a linear speech stream: the influence of morphological structure on the recognition of spoken Arabic words. Brain Lang. 147, 1–13. https://doi.org/10.1016/j.bandl.2015.04.006.
Hale, J., 2006. Uncertainty about the rest of the sentence. Cogn. Sci. 30 (4), 643–672. https://doi.org/10.1207/s15516709cog0000_64.
Hale, J., 2016. Information-theoretical complexity metrics. Lang. Linguist. Compass 10 (9), 397–412. https://doi.org/10.1111/lnc3.12196.
Henderson, J.M., Choi, W., Lowder, M.W., Ferreira, F., 2016. Language structure in the brain: a fixation-related fMRI study of syntactic surprisal in reading. Neuroimage 132, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050.
Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? Lang. Cognit. Neurosci. 31 (1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.
Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307 (5947), 161. https://doi.org/10.1038/307161a0.
Lau, E., Stroud, C., Plesch, S., Phillips, C., 2006. The role of structural prediction in rapid syntactic analysis. Brain Lang. 98 (1), 74–88. https://doi.org/10.1016/j.bandl.2006.02.003.
Lau, E.F., Holcomb, P.J., Kuperberg, G.R., 2013. Dissociating N400 effects of prediction from association in single-word contexts. J. Cogn. Neurosci. 25 (3), 484–502. https://doi.org/10.1162/jocn_a_00328.
Lau, E.F., Weber, K., Gramfort, A., Hämäläinen, M.S., Kuperberg, G.R., 2016. Spatiotemporal signatures of lexical–semantic prediction. Cerebr. Cortex 26 (4), 1377–1387. https://doi.org/10.1093/cercor/bhu219.
Linzen, T., Jaeger, T.F., 2015. Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. Cogn. Sci. 40 (6), 1382–1411. https://doi.org/10.1111/cogs.12274.
Maess, B., Schröger, E., Widmann, A., 2016. High-pass filters and baseline correction in M/EEG analysis. Commentary on:"How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition". J. Neurosci. Methods 266, 164–165.
Maess, B., Mamashli, F., Obleser, J., Helle, L., Friederici, A.D., 2016. Prediction signatures in the brain: semantic pre-activation during language comprehension. Front. Hum. Neurosci. 10, 591. https://doi.org/10.3389/fnhum.2016.00591.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods 164 (1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.
Matchin, W., Hammerly, C., Lau, E., 2017. The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. Cortex 88, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010.
Molinaro, N., Barber, H.A., Carreiras, M., 2011. Grammatical agreement processing in reading: ERP findings and future directions. Cortex 47 (8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019.
Nelson, M., Dehaene, S., Pallier, C., Hale, J., 2017. Entropy Reduction correlates with temporal lobe activity. In: Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2017), pp. 1–10.
Nelson, M.J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S.S., Naccache, L., Hale, J.T., Pallier, C., Dehaene, S., 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. Proc. Natl. Acad. Sci. 114 (18), E3669–E3678. https://doi.org/10.1073/pnas.1701590114.
Neville, H., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F., 1991. Syntactically based sentence processing classes: evidence from event-related brain potentials. J. Cogn. Neurosci. 3 (2), 151–165. https://doi.org/10.1162/jocn.1991.3.2.151.
Nieuwland, M.S., 2019. Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. Neurosci. Biobehav. Rev. 96, 367–400. https://doi.org/10.1016/j.neubiorev.2018.11.019.
Pattamadilok, C., Dehaene, S., Pallier, C., 2016. A role for left inferior frontal and posterior superior temporal cortex in extracting a syntactic tree from a sentence. Cortex 75, 44–55. https://doi.org/10.1016/j.cortex.2015.11.012.
Peirce, J.W., 2007. PsychoPy—psychophysics software in Python. J. Neurosci. Methods 162 (1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017.
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72 (2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6.
Pylkkänen, L., Bemis, D.K., Elorrieta, E.B., 2014. Building phrases in language production: an MEG study of simple composition. Cognition 133 (2), 371–384. https://doi.org/10.1016/j.cognition.2014.07.001.
Saiegh-Haddad, E., Henkin-Roitfarb, R., 2014. The structure of Arabic language and orthography. In: Handbook of Arabic Literacy. Springer, Dordrecht, pp. 3–28.
Sassenhagen, J., Draschkow, D., 2019. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology 56 (6), e13335. https://doi.org/10.1111/psyp.13335.
Schiffer, A.M., Ahlheim, C., Wurm, M.F., Schubotz, R.I., 2012. Surprised at all the entropy: hippocampal, caudate and midbrain contributions to learning from prediction errors. PLoS One 7 (5). https://doi.org/10.1371/journal.pone.0036445.
Solomyak, O., Marantz, A., 2010. Evidence for early morphological decomposition in visual word recognition. J. Cogn. Neurosci. 22 (9), 2042–2057. https://doi.org/10.1162/jocn.2009.21296.
Speer, R., Chin, J., Lin, A., Jewett, S., Nathan, L., 2017, September 27. . LuminosoInsight/wordfreq: v1.7. Zenodo.
Taft, M., 2004. Morphological decomposition and the reverse base frequency effect. Q. J. Exp. Psychol. Sect. A. 57 (4), 745–765. https://doi.org/10.1080/02724980343000477.
Tanner, D., Morgan-Short, K., Luck, S.J., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology 52 (8), 997–1009. https://doi.org/10.1111/psyp.12437.
Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L., Salmelin, R., 1999. Dynamics of letter string perception in the human occipitotemporal cortex. Brain 122 (11), 2119–2132. https://doi.org/10.1093/brain/122.11.2119.
van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol. Learn. Mem. Cogn. 31 (3), 443. https://doi.org/10.1037/0278-7393.31.3.443.
Van Heuven, W.J., Mandera, P., Keuleers, E., Brysbaert, M., 2014. SUBTLEX-UK: a new and improved word frequency database for British English. Q. J. Exp. Psychol. 67 (6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521.
Warren, R.M., 1970. Perceptual restoration of missing speech sounds. Science 167 (3917), 392–393. https://science.sciencemag.org/content/167/3917/392.
Westerlund, M., Kastner, I., Al Kaabi, M., Pylkkänen, L., 2015. The LATL as locus of composition: MEG evidence from English and Arabic. Brain Lang. 141, 124–134. https://doi.org/10.1016/j.bandl.2014.12.003.
Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., Van den Bosch, A., 2016. Prediction during natural language comprehension. Cerebr. Cortex 26 (6), 2506–2516. https://doi.org/10.1093/cercor/bhv075.
Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., Paavilainen, P., 2016. Predictive coding of phonological rules in auditory cortex: a mismatch negativity study. Brain Lang. 162, 72–80. https://doi.org/10.1016/j.bandl.2016.08.007.
Zaccarella, E., Friederici, A.D., 2015. Merge in the human brain: a sub-region based functional investigation in the left pars opercularis. Front. Psychol. 6, 1818. https://doi.org/10.3389/fpsyg.2015.01818.
Zaccarella, E., Schell, M., Friederici, A.D., 2017. Reviewing the functional basis of the syntactic Merge mechanism for language: a coordinate-based activation likelihood estimation meta-analysis. Neurosci. Biobehav. Rev. 80, 646–656. https://doi.org/10.1016/j.neubiorev.2017.06.011.