Electrophysiological evidence for an independent effect of memory retrieval on referential processing

Hossein Karimi[a,*], Tamara Y. Swaab[b], Fernanda Ferreira[b]

[a] *Pennsylvania State University*
[b] *University of California, Davis*

ABSTRACT

In this study, we show that the difficulty of re-activating and retrieving the representations of potential referents from memory (retrieval difficulty) influences referential processing, and that this effect is independent of the number of potential referents for a pronoun or the probability of possible referential interpretations (referential coherence). In two experiments, we varied retrieval difficulty by manipulating whether two referential candidates were modified by extra semantic information or not, creating representationally rich (modified) or bare (unmodified) referential candidates, respectively, and we measured event-related brain potentials (ERPs) on following pronouns. We observed a sustained frontal negative shift (the Nref effect) on pronouns following bare, and therefore difficult-to-retrieve, referential candidates relative to those following representationally rich candidates, regardless of the ambiguity of pronouns and the probability of either referential interpretation. Since referential coherence was held constant across the conditions, the results suggest that retrieval difficulty affects referential processing independently of coherence. We discuss the implications for memory-based theories of language processing.

Introduction

Pronouns such as *he* and *she* are commonly used in everyday language. Normally, pronouns carry little semantic information and derive their meaning from the entities they refer to, that is, their referents. Thus, processing a pronoun involves identifying a unique referent for it. For example, in *The actor had a long walk with the actress around the lake. He seemed a little nervous*, processing the pronoun *he* would in most cases result in co-indexation of *he* with the first-mentioned Noun Phrase (henceforth NP, i.e., *the actor*).

In this study, we draw a distinction between two potential sources of difficulty associated with resolving a referential dependency such as the one illustrated above. First, the difficulty of establishing a referential link could vary depending on the number of referential candidates that could potentially serve as the referent for a referring expression, and/or the probability of possible referential interpretations. Specifically, multiple or no potential referents could lead to more processing difficulty compared to a situation in which there is a single clear referent. Similarly, equi-probable referential interpretations could result in more processing difficulty than cases in which one interpretation is more likely than the other(s). We will henceforth call this property *referential coherence* (Nieuwland & Martin, 2017) because having one straightforward and unique referential interpretation results

in a more coherent discourse compared to discourses in which multiple interpretations are possible or equally likely. The second source of difficulty associated with resolving a referential dependency might be based on how easily the memory representation associated with a referent can be retrieved from memory, as referential processing necessarily involves encoding a referential candidate in memory and then reactivating and retrieving that representation when a pronoun (or any other referring expression) is encountered (Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; Gerrig & McKoon, 1998; Lucas, Tanenhaus, & Carlson, 1990; MacDonald & MacWhinney, 1990; Sanford & Garrod, 1989, 2005). We will henceforth refer to this difficulty as *retrieval difficulty* as it captures the difficulty associated with retrieving the relevant representations from memory. The logic behind this distinction between referential coherence and retrieval difficulty is that even when a referring expression is coherent—that is, appears to have a single referent—the representation associated with that referent could be easy or difficult to retrieve depending on how activated it is in memory.

What happens when a referring expression has multiple rather than a single potential referent, and the discourse is consequently less coherent? Numerous studies have demonstrated that when multiple potential interpretations are plausible, the referential candidate that is relatively more activated in memory is retrieved faster and is taken to be the referent. For instance, it has been repeatedly shown that the

syntactic subject of a sentence is considerably more likely to be interpreted as co-referential with an ambiguous pronoun compared to the syntactic object (e.g., Arnold, 2001; Fletcher, 1984; Fukumura & Van Gompel, 2010; Givón, 1983; Gordon, Grosz, & Gilliom, 1993; Gundel, Hedberg, & Zacharski, 1993; Karimi & Ferreira, 2016a), presumably because the subject role confers more prominence on the associated NP (Gordon et al., 1993; Gundel et al., 1993), increasing its baseline activation level (e.g., Jäger, Engelmann, & Vasishth, 2017). The greater prominence of the syntactic subject is also consistent with the fact that it usually coincides with the "topic" of a sentence in English and topicality has also been shown to contribute to prominence (Ariel, 1990; Gundel et al., 1993; Rohde & Kehler, 2014). Similarly, ambiguous pronouns have been shown to be interpreted as referring to animate rather than inanimate NPs (Fukumura & Van Gompel, 2011) as well as to the semantically richer NPs (Karimi & Ferreira, 2016a). Since both animacy and semantic richness have been associated with greater activation levels in memory (animacy: Bock & Warren, 1985; Branigan, Pickering, & Tanaka, 2008; Rosenbach, 2008; semantic richness: Hofmeister, 2011; Troyer, Hofmeister, & Kutas, 2016), these results suggest that the referential candidate that is relatively more activated is taken to be the referent of an ambiguous pronoun. Thus, the relative activation levels of referential candidates directly influence discourse coherence by determining "what" (i.e., which referential candidate) is retrieved and taken to be the referent of the pronoun. In other words, when more than one referential interpretation is possible, the probability of a specific interpretation depends on the relative activation levels of the referential candidates, with the ultimate interpretation being consistent with the more easily retrievable candidate.

Numerous studies have investigated referential processing by recording event-related brain activity from the scalp (event-related potentials or ERPs). An effect of particular relevance to the present study is the *Nref* effect– a late and sustained frontal negativity that emerges in one condition relative to another and reflects referential processing difficulty[1] (e.g., Martin, Nieuwland, & Carreiras, 2012, 2014; Nieuwland, 2014; Nieuwland, Otten, & Van Berkum, 2007; Nieuwland & Van Berkum, 2006; Nieuwland & Van Berkum, 2008; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). The difficulty that has been observed in these studies may be attributable either to referential coherence, retrieval difficulty, or both. To test for retrieval difficulty it is necessary to hold constant all known factors that could contribute to referential coherence, and this is the major goal of the current study. In the remainder of this section, we first discuss the role of memory retrieval during referential processing, and then we will review relevant findings from previous studies in relation to our research question, arguing that current findings are equivocal with regards to the question whether ease of retrieval or referential coherence influence referential processing.

The role of memory retrieval during referential processing

Referential processing necessarily involves reactivating and retrieving the memory representations associated with the referential candidates from memory when a referring expression is encountered (Dell et al., 1983; Gernsbacher, 1989; Gerrig & McKoon, 1998; Lucas et al., 1990; MacDonald & MacWhinney, 1990; Sanford & Garrod, 1989, 2005). Therefore, one would expect ease of reactivation and retrieval from memory to influence referential processing and hence elicit an Nref effect (see Martin et al., 2012, 2014). This possibility is in line with cue-based retrieval theories of language processing. According to these

theories, resolving a linguistic dependency such as the one between a pronoun and its referent involves a content-addressable memory search in which items needed for resolving the dependency are directly accessed via a cue-based retrieval mechanism (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2009; McElree, 2006; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2006; also see Jäger et al., 2017).

Under these theories, a word whose processing depends on some previously encoded item is assumed to trigger the retrieval of that item for successful dependency formation. Crucially, the ease with which the target memory representation is retrieved is assumed to determine the difficulty of establishing the linguistic dependency (also see Gibson, 1998, 2000). In the original versions of cue-based retrieval theories, the retrieval of an item from memory is a function of its baseline activation level as well as the amount of interference that is experienced during its retrieval (e.g., Lewis & Vasishth, 2005; Lewis et al., 2006). Relevant to the current study, Jäger et al. (2017) performed a meta-analysis of current findings related to processing syntactic and referential dependencies and identified "competitor prominence" as an additional factor that influences the retrieval of the target item: The more prominent a competitor, the more difficult the retrieval of the intended (i.e., target) item. In the context of referential processing, cue-based retrieval predicts that the morpho-syntactic features of a pronoun such as its number and gender. (i.e., the retrieval cues) should activate the representations of all the referents that match those features to varying levels depending on the degree of match such that the representation that matches the most features should be activated the most and should eventually win the race for retrieval. Importantly, then, the ease with which the target representation (i.e., the referent) can be retrieved (determined by baseline activation, retrieval interference and competitor prominence) should affect referential processing difficulty and, therefore, the Nref.

Retrieval difficulty versus referential coherence

Previous studies have demonstrated that the Nref effect is elicited for less referentially coherent referring expressions relative to more coherent ones, suggesting that less coherent referring expressions are more difficult to process. However, these results are agnostic with respect to the distinction between retrieval difficulty and referential coherence. Broadly speaking, referential coherence has been manipulated via three factors in the current literature on referential processing: ambiguity, contextual bias in favor of possible referential interpretations, and feature mismatch between a pronoun and a single preceding referential candidate. Note that ambiguity and feature mismatch influence the number of potential referents for a pronoun, but contextual bias influences the probability of possible referential interpretations. Specifically, ambiguous referring expressions have been demonstrated to result in an Nref effect compared to unambiguous referring expressions, which is taken to reflect greater processing difficulty associated with referential ambiguity (e.g., Nieuwland & Van Berkum, 2006; Van Berkum et al., 1999 see also Nieuwland, Petersson, & Van Berkum, 2007; Nieuwland et al., 2007; Van Berkum et al., 2003).

Similarly, controlling for ambiguity, Nieuwland and Van Berkum (2006) manipulated verb bias and reported a greater Nref effect on ambiguous pronouns following contexts that provided weak support for either of two referential interpretations (e.g., *John hit David while he …*) compared to contexts that strongly supported one interpretation (e.g., *John hated David because he …*). In other words, referential processing was found to be easier when one referential candidate was more likely than the other to occur next given the preceding discourse. This is presumably because although the number of potential referents is formally the same, a strong bias makes one referential candidate much more probable as the referent than the other, leading to more coherent discourse than in contexts in which both referential interpretations are equally likely.

---

[1] The difficulty of referential processing has also been shown to result in N400 and P600 modulations (e.g., Swaab, Camblin, & Gordon, 2004; Ledoux, Gordon, Camblin, & Swaab, 2007; Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Johns, Gordon, Long, & Swaab, 2014; Lamers, Jansma, Hammer, & Munte, 2006). However, since these studies were concerned with other aspects of referential processing such as the structural prominence of referents or the effects of referential failure, we will not discuss them here.