



Humanwissenschaftliche Fakultät

Sol Lago | Anna Namyst | Lena A. Jäger | Ellen Lau





Antecedent access mechanisms in pronoun processing

evidence from the N400

Suggested citation referring to the original publication:
Language, Cognition and Neuroscience 34 (2019) 5, pp. 641-661
DOI <https://doi.org/10.1080/23273798.2019.1566561>
ISSN (print) 2327-3798
ISSN (online) 2327-3801

Postprint archived at the Institutional Repository of the Potsdam University in:
Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 568
ISSN 1866-8364
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-433237>
DOI <https://doi.org/10.25932/publishup-43323>

Antecedent access mechanisms in pronoun processing: evidence from the N400

Sol Lago ^a, Anna Namyst ^b, Lena A. Jäger ^a and Ellen Lau ^b

^aDepartment of Linguistics, University of Potsdam, Potsdam, Germany; ^bDepartment of Linguistics, University of Maryland, College Park, MD, USA

ABSTRACT

Previous cross-modal priming studies showed that lexical decisions to words after a pronoun were facilitated when these words were semantically related to the pronoun's antecedent. These studies suggested that semantic priming effectively measured antecedent retrieval during coreference. We examined whether these effects extended to implicit reading comprehension using the N400 response. The results of three experiments did not yield strong evidence of semantic facilitation due to coreference. Further, the comparison with two additional experiments showed that N400 facilitation effects were reduced in sentences (vs. word pair paradigms) and were modulated by the case morphology of the prime word. We propose that priming effects in cross-modal experiments may have resulted from task-related strategies. More generally, the impact of sentence context and morphological information on priming effects suggests that they may depend on the extent to which the upcoming input is predicted, rather than automatic spreading activation between semantically related words.

ARTICLE HISTORY

Received 7 June 2018
Accepted 29 December 2018

KEYWORDS

Coreference; semantic priming; event-related potentials; sentence comprehension; N400

Introduction


Accurate sentence interpretation often depends on the correct treatment of pronouns. For example, speakers are likely to judge a sentence as semantically coherent if the preamble *Susan always knew that the bachelor loved his ...* is followed by words like *freedom* or *family*, but certainly not *wife*. The unsuitability of *wife* can be detected because speakers link the pronoun to the antecedent *the bachelor* and retrieve its semantic features, which indicate that bachelors are unmarried individuals. Thus, the speed of antecedent retrieval will determine how quickly properties of the sentence and the discourse can be computed. The current work uses event-related potentials to investigate the mechanisms underlying this retrieval process in comprehension.

A major challenge to study antecedent retrieval during coreference is to locate an early and implicit measurement that indicates whether an antecedent has been successfully retrieved. Previous studies have tried to diagnose antecedent retrieval through the presence of *semantic relatedness* or *priming effects* – evidence of facilitated processing for words that follow a pronoun and that are semantically associated with its antecedent. These effects could in principle arise due to several processing mechanisms. First, since the representation of the antecedent's *discourse referent* contains its

conceptual features, its reactivation in speakers' working memory of the discourse could result in spreading activation to semantically related concepts (Cloitre & Bever, 1988; Lucas, Tanenhaus, & Carlson, 1990). Second, encountering a pronoun might additionally result in the reactivation of its *lexical antecedent* in speakers' long-term memory (e.g. the word *bachelors* in the lexicon). The reactivation of this word in the lexicon should result in spreading activation to semantically related words, according to accounts in which semantically associated words are stored close together or share heavily-weighted connections (Collins & Loftus, 1975; Forster, 1976; Levelt, Roelofs, & Meyer, 1999; Morton, 1979).

Seminal work with cross-modal lexical decision in English proposed that semantic association effects occurred early and that they could be used as an effective measure of antecedent reactivation (Leiman, 1982; Nicol, 1988; Shillcock, 1982). In these studies, participants would hear a sentence with a pronoun and would perform a visual lexical decision task (the asterisks below mark the points in which the visual probe appeared in different trials). The target of the lexical decision was a semantically related word (e.g. *books*), or a semantically unrelated word that was matched in length and frequency (e.g. *trees*):

CONTACT Sol Lago  marlago@uni-potsdam.de

 Supplemental data for this article can be accessed here [10.1080/23273798.2019.1566561](https://doi.org/10.1080/23273798.2019.1566561).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The motorbike could not be returned to the **author** * as was originally planned, principally * because at the time **he** * was in the South of France.

Cross-modal studies reported that immediately after the pronoun's offset, participants made faster lexical decisions to semantically related words, consistent with processing facilitation. By contrast, when the pronoun was replaced by a non-coreferential form, such as *it*, lexical decision times did not differ between related and unrelated words. This suggested that the previously observed facilitation was not due to residual activation from the initial occurrence of the antecedent noun, but rather by its reactivation specifically due to coreference.

However, the conclusions that can be drawn from cross-modal priming paradigms have several limitations. First, cross-modal paradigms may engage conscious strategic computations, as detecting semantic relationships between words improves participants' performance in the lexical decision task, which might encourage them to focus on semantic relationships to perform better (Neely, 1991). As a result, priming effects in cross-modal studies may be partly attributable to task-related strategies, rather than automatic antecedent reactivation. Second, it is difficult to estimate the time course of antecedent reactivation from lexical decision responses, which measure keypress latencies to visual probes, typically around 600–800 ms post-probe onset. These latencies provide very rough timing measures, because they simultaneously reflect the effect of antecedent reactivation and the delay associated with manual motor responses.

More recently, work using eye-tracking during reading has cast some doubt on whether early semantic effects occur during implicit comprehension, at least in languages like English (Lago et al., 2017). Lago and colleagues examined whether priming effects varied as a function of the grammatical properties of a language. They compared German, a language with syntactic gender (masculine, feminine or neuter) with English, a language in which gender is only conceptual (male, female). They hypothesised that antecedent retrieval in German might include reactivation of both the discourse referent and the pronoun's antecedent in the lexicon, where grammatical gender is stored (Cacciari, Carreiras, & Barbolini-Cionini, 1997; Frazier, Henstra, & Flores d' Arcais, 1996; Garnham, Oakhill, Erlich, & Carreiras, 1995). By contrast, the reactivation of a lexical antecedent may not be needed in English, as the retrieval of a discourse referent should be sufficient to license the pronoun's features, such as conceptual gender, animacy and number. To address

this hypothesis, the study implemented semantic relatedness paradigms using English and German passages with possessive pronouns. Crucially, the passages varied whether the antecedents were semantically related or unrelated to the word after the pronoun:

The maintenance men told the **singer**^{RELATED} / **deputy**^{UNRELATED} about a problem.

They had broken **his** piano and would have to repair that first.

Across experiments, English and German speakers showed reading facilitation for the target word, *piano*, when it was preceded by a semantically related antecedent. But crucially, whereas both groups showed facilitation in late reading measures (i.e. re-reading and total reading times), only German speakers showed facilitation in early measures (i.e. single fixation and first fixation times) immediately after the pronoun. The authors proposed that the lack of early facilitation effects in English was due to the sole reactivation of the discourse referent, whereas in German both the referent and the lexical representation of the antecedent were reactivated in order to license the grammatical gender of the pronoun. They suggested that whereas lexical reactivation resulted in spreading activation to semantically associated words, the reactivation of the discourse referent did not, thus failing to pre-activate them. By contrast, facilitation effects in late measures, which were observed across languages, were proposed to index sentence-level integration processes that were likely occur well after lexical and discourse antecedent reactivation.

If this previous account is correct, it suggests that semantic relatedness effects may not serve as a measure of antecedent reactivation cross-linguistically, because they result from a lexical retrieval process that is engaged only in languages with grammatical gender. However, there are several reasons for caution. First, the absence of early relatedness effects in English contrasts with the earlier cross-modal priming results (although as noted above, these could also be attributed to strategic processes not characteristic of normal comprehension). Second, semantic relatedness effects in previous English eye-tracking studies have often shown weak or unreliable effects on the eye-movement record, at least in studies that have compared them with discourse-level variables, such as information structure and plausibility (Camblin, Gordon, & Swaab, 2007; Morris, 1994; Traxler, Foss, Seely, Kaup, & Morris, 2000). Therefore, it is possible that the lack of relatedness effects in previous English eye-tracking work was partly due to properties of the paradigm used.

In sum, semantic relatedness effects could serve as a powerful tool for investigating the antecedent retrieval mechanisms that support pronoun interpretation, but prior evidence is mixed as to whether these effects can reliably reflect antecedent reactivation in languages like English, which lack grammatical gender. In the present study, we sought to resolve this conflict by using a different measure of comprehension, event-related potentials (ERPs). ERPs are ideally suited to examine this question because they have high temporal resolution, they are well-established to be sensitive to semantic relatedness and they do not require participants to make conscious decisions about the semantic relationships under investigation.

The present study

The ERP experiments presented below examined the question of whether English pronouns rapidly reactivate a semantic representation of their antecedents during coreference. We constructed passages with possessive expressions that consisted of repeated noun phrases (henceforth repeated noun phrases or NPs, e.g. *the prince's* in 1a,b) and pronouns (*his/her* in 1c,d,e). Similarly to previous studies, we varied the relationship between the word after the possessive expression, such that it could be semantically related (*castle* in 1a,c) or unrelated (*estate* in 1b,d,e) to the possessive's antecedent:

- (1) *Preamble*: Genevieve travelled with the prince for many weeks.
 - (a) *Repeated NP, related*: She wished that the **prince's** castle wasn't so far away.
 - (b) *Repeated NP, unrelated*: She wished that the **prince's** estate wasn't so far away.
 - (c) *Pronoun, related*: She wished that **his** castle wasn't so far away.
 - (d) *Pronoun, unrelated*: She wished that **his** estate wasn't so far away.
 - (e) *Coreference control*: She wished that **her** castle wasn't so far away.

In order to assess the presence of semantic facilitation effects, we measured brain responses to the word after the possessive (henceforth the *target word*) using an ERP component known as the N400. The N400 is a negative deflection that peaks around N400 ms and is typically maximal over centro-parietal electrodes (for review, see Kutas & Federmeier, 2011). Importantly, the N400 component is highly sensitive to semantic relatedness, such that when a word is preceded by a semantically related word in lexical decision tasks, its N400

response is reduced (less negative), than when it is preceded by a semantically unrelated word (Brown, Hagoort, & Chwilla, 2000; Holcomb, 1988; Kutas & Van Petten, 1988, 1994; Rugg, 1985).

Our experimental predictions were as follows. With repeated NPs (1a,b, *the prince's*), we expected the preceding possessive NP to prime the semantically related target word *castle*, yielding a smaller N400 response than for the unrelated word *estate*. In the critical conditions, the repeated NP was replaced by a coreferential pronoun (1c,d, *his*). We hypothesised that if comprehenders rapidly reactivated the semantic features of the antecedent upon encountering the pronoun, N400 responses to semantically related words should yield an N400 modulation similar to that in the repeated NP conditions. In addition, we sought to ensure that facilitation to the target word *castle* was not due to residual activation from having read the NP *the prince* in the context sentence. This was achieved by the coreference control condition (1e, *her*), which kept the context containing the related target word constant but used a pronoun that no longer referred to the semantically related antecedent, but to the proper name antecedent (e.g. *Genevieve*). We predicted that if semantic facilitation was specifically due to coreference rather than a more general context-priming effect, then responses in the coreference control condition (1e) should not be facilitated as compared with (1c).

The ERP data was analyzed using Bayesian models, because they offer several advantages over frequentist models (Gelman et al., 2014; McElreath, 2015). One limitation of frequentist statistics is that they do not provide direct information about the alternative to the null hypothesis (which is the hypothesis that a given effect is indeed absent), even though this alternative is often the hypothesis of interest. Rather, a frequentist *p*-value describes the probability of the observed data given the null hypothesis. When the observed data is deemed very unlikely, the null hypothesis can be rejected and thus the alternative hypothesis is indirectly supported. Notably, the rejection of the null hypothesis does not afford any conclusions about the size of the effects of interest. But in modern psycholinguistics, obtaining realistic estimates of effect sizes is crucial, as computational models need not only qualitative information about the presence or absence of an effect, but also accurate quantitative estimates that allow the evaluation of the model's predictions, such that theory building, modelling and empirical work can mutually inform one another. For the present study, the importance of precise effect size estimates was particularly relevant given previous eye-tracking findings that priming

effects due to the antecedent reactivation were very small in English comprehension (Lago et al., 2017).

The Bayesian paradigm provided the statistical tools for the calculation of precise effect estimates and the uncertainty associated with them. There are three notions that are central in this paradigm: the *posterior*, *likelihood* and *prior*. The *posterior* represents a probability distribution over different model parameters given the data and the model. Thus, the marginalised posterior distribution of an effect can be intuitively understood as the probability of different effect sizes given the data. In contrast, the *likelihood* defines the distribution that is believed to underlie the generation of the data, with the parameters of this distribution representing the entities of interest to be estimated. Finally, the *prior* distribution reflects the researcher's prior belief about how likely different parameter values are. Priors can be either uninformative or regularising. Uninformative priors make all parameter values equally likely, and thus the prior does not strongly influence the posterior. By contrast, regularising priors are more informative and they do make assumptions about the distribution of the parameter values, thus ensuring that whenever there is not enough data to support an effect, the posterior tends towards a conservative estimate.

The computation of Bayesian statistics was performed with hierarchical linear models, also known as linear-mixed models. We used random intercepts and slopes to account for the variance between items and participants. This approach allowed us to avoid the aggregation of ERP data over items or subjects, thus increasing statistical power. For research using hierarchical linear models for the analysis of ERP data within the frequentist paradigm see previous work by Franck and colleagues (Frank, Otten, Galli, & Vigliocco, 2015) and Dambacher and colleagues (Dambacher, Kliegl, Hofmann, & Jacobs, 2006). Bayesian hierarchical models have been used to analyze ERP data in recent work by Nieuwland and colleagues (Nieuwland et al., 2018).

Overview of the studies

The experiments were organised as follows. Experiment 1 tested materials such as the ones shown in (1). Experiments 2 and 3 were direct replications of Experiment 1 but split the 5-condition design into two separate studies to increase the number of trials per participant and condition. Experiment 2 focused on the repeated NP conditions (1a,b), and Experiment 3 focused on the pronoun conditions (1c,d,e). Experiments 4 and 5 addressed the existence of priming effects in single-word paradigms. Experiment 4 presented the word pairs as bare nouns (*prince*–*castle* vs. *prince*–*estate*) and

Experiment 5 presented the prime words with possessive case markers, rendering them more similar to the phrasal stimuli in the sentence experiments (*prince's*–*castle* vs. *prince's*–*estate*).

Experiment 1

Experiment 1 tested 5-condition item sets such as (1). Based on previous findings that N400 responses to a word are facilitated when it is preceded by a semantically unrelated word (Brown et al., 2000; Holcomb, 1988; Kutas & Van Petten, 1994; Rugg, 1985), we expected priming for words that were preceded by a semantically related possessive NP (1a vs. 1b). Our research question was whether the same priming effect would occur when the preceding NP was replaced by a coreferential pronoun (1c vs. 1d). Further, if the priming effect was specifically due to coreference, rather than to residual activation from the NP *the prince* in the previous context, then responses in the coreference control condition (1e) should not be facilitated as compared with (1c).

Methods

Participants

Participants in this and following ERP studies were right-handed adult native speakers of American English recruited from the University of Maryland. Participants received monetary compensation or course credit for their contribution and gave informed consent in accordance with the Institutional Review Board of the University of Maryland. Thirty-seven subjects participated in Experiment 1, but seven were excluded from analysis due to excessive artifacts (more than 40% of trials rejected). Thirty participants were included in the final analysis (mean age = 22 years, age range = 18–30 years, 19 females).

Materials

We created 150 experimental passages distributed across five conditions, as shown in (1). The first sentence in the passage was identical across conditions and introduced two characters as the subject and object of the main verb. The object was always a noun with a clear gender bias (e.g. *prince*), while the subject was a proper name biased towards a different gender (e.g. *Genevieve*). The two characters always differed in gender to ensure that only one was a suitable antecedent for the possessive expression in the following sentence. The second sentence manipulated whether the possessive expression was a repeated noun phrase or a pronoun and whether the word immediately after the possessive expression was semantically related or unrelated to the

target antecedent. In the remaining condition, the pronoun differed in gender from the antecedent noun, thus referring unambiguously to the proper noun.

To select the antecedent nouns, we began by choosing nouns describing strongly gender-biased occupations or roles, such as tailor or prince. These nouns were normed for gender bias by twenty participants recruited on Amazon Mechanical Turk using a 1–7 scale counterbalanced by list and standardised for analysis (for details, see Chow, Lewis, & Phillips, 2014). Antecedent nouns had a clear gender bias (i.e. rated lower than 3 or greater than 5) or were combined with a stereotyped modifier to bias the reader towards the intended gender (e.g. *the pregnant instructor*, *the bearded manager*). We note that more of the occupations were gender-stereotyped towards male than female referents (86% vs. 14%), and so the proportion of male and female pronouns in the pronoun conditions were correspondingly asymmetrical.

For each antecedent noun, we chose a semantically related word (e.g. *castle*) using the South Florida Free Word Association Norms as a guide (Nelson, McEvoy, & Schreiber, 1998). We avoided target words that overlapped orthographically and phonologically with the antecedent noun and highly frequent words that could elicit floor effects in the N400 response (Van Petten & Kutas, 1990). For each semantically related target word, we then used the English Lexicon Project database (Balota et al., 2007) to select an unrelated word that was matched to the related noun on word length ($\text{mean}_{\text{RELATED}} = 6.30$; $\text{SD}_{\text{RELATED}} = 1.95$; $\text{mean}_{\text{UNRELATED}} = 6.36$; $\text{SD}_{\text{UNRELATED}} = 1.87$) and log word frequency ($\text{mean}_{\text{RELATED}} = 2.86$; $\text{SD}_{\text{RELATED}} = 0.66$; $\text{mean}_{\text{UNRELATED}} = 2.85$; $\text{SD}_{\text{UNRELATED}} = 0.66$), as determined by the SUBTLEXus database (Brysbaert & New, 2009). Finally, the semantic association between related and unrelated target words and the antecedent nouns was quantified by forty-two participants from the University of Maryland using a 1–5 (“not related”– “very related”) scale. The results indicated a large difference in ratings between conditions: related pairs had a mean rating of 4.54 ($\text{SD} = 0.46$), whereas unrelated pairs had a mean rating of 2.05 ($\text{SD} = 0.74$).

In addition to the semantic association norming, we also measured the cloze probability of related and unrelated target words, as it is known that predictability can impact the N400 response (e.g. Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas, 1993). Eighty participants recruited on Amazon Mechanical Turk were asked to complete the experimental passages truncated after the possessive expression (split evenly between full NP possessives and the corresponding possessive pronoun). Each presentation list contained half of the items, resulting in 40 completion responses per item.

The results indicated that the sentence contexts did not strongly or consistently predict related targets. The mean cloze value for related target words after gender-matching possessive expressions was 5% (for words after pronouns, $\text{SD} = 12\%$) and 5% (for words after repeated NPs, $\text{SD} = 12\%$). By contrast, the mean cloze value for unrelated target words after gender-matching possessive expressions was 0.7% (for words after pronouns, $\text{SD} = 2\%$) and 0.6% (for words after repeated NPs, $\text{SD} = 2\%$). Approximately half the items had a cloze of 0.

Experimental items were distributed across 5 lists in a Latin Square design, such that each list contained exactly one version of each item and 30 items per condition. Each participant saw exactly one version of each item. The experiment also contained 72 two-sentence filler items of comparable length and complexity. Filler items contained other kinds of referential expressions and anaphors, particularly feminine pronouns to balance the larger proportion of masculine pronouns in the experimental items.

Procedure

Participants were seated in a dimly lit room. Stimuli were visually presented one word at a time on a computer monitor in white 24-point case Arial font on a black background. Each word appeared for 300 ms with an interstimulus interval of 200 ms ($\text{SOA} = 500$ ms). The experimental session was divided into 5 blocks separated by rest intervals. A third of the trials was followed by yes/no comprehension questions, which ensured that participants were attending to the stimuli. The questions never alluded to the referential dependency, to avoid focusing participants' attention on this relationship. Half of the questions asked about the first sentence and half asked about the second sentence. Target yes/no answers were in a 1:1 ratio. The order of experimental and filler items was randomised for each list. Experimental sessions lasted approximately 60 min, in addition to preparation and clean up time.

Electrophysiological recording

Twenty-nine tin electrodes were held on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH) in a 10–20 configuration (O1, Oz, O2, P7, P3, Pz, P4, P8, TP7, Cp3, CPz, CP4, TP8, T7, C3, Cz, C4, T8, FT7, FC3, FCz, FC4, FT8, F7, F3, Fz, F4, F8, FP1). Bipolar electrodes were placed above and below the left eye and at the outer canthus of the right and left eyes to monitor vertical and horizontal eye movements. Additional electrodes were placed over the left and right mastoids. Scalp electrodes were referenced online to the left mastoid and re-referenced offline to the average of left and right

mastoids. Impedances were maintained at less than 10 k Ω for all scalp electrode sites and less than 5 k Ω for mastoid sites and ocular electrodes. The EEG signal was amplified by a NeuroScan SynAmps[®] Model 5083 (NeuroScan, Inc., Charlotte, NC) with a bandpass of 0.05–100 Hz and was continuously sampled at 500 Hz by an analog-to-digital converter.

Analysis

Averaged ERPs time-locked to the possessive expression were formed offline from trials free of ocular and muscular artifact using preprocessing routines from the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes. 15.39% of trials were rejected due to excessive artifacts (rejections range: 1%–34%). Epochs with artifacts were excluded, and channels with a disproportionate number of epochs containing peak-to-peak fluctuations of 100 μ V or more were interpolated. For the nine channels of interest, this procedure affected only one channel in one participant. A 100-ms pre-stimulus baseline was subtracted from all waveforms before statistical analysis, and a 40-Hz low-pass filter was applied to the ERPs offline. ERP data for this and following experiments, as well as all experimental materials, are publicly available in the Open Science Framework repository (<https://osf.io/>). All analyses were conducted on mean ERP amplitudes for the target words in the 300–500 ms time-window, across a set of 9 centro-parietal electrodes in which the N400 effect is typically maximal: P3, Pz, P4, CP3, CPz, CP4, C3, Cz, C4. The time-window and electrodes of interest were kept constant across experiments to avoid selection bias.

Statistical analyses were conducted with R (R Development Core Team, 2018). Bayesian analyses were performed with the package *rstan* (Stan Development Team, 2017a), which makes use of the probabilistic programming language Stan (Carpenter et al., 2017), and with the package *brms* (Bürkner, 2016). After aggregating the data over the electrodes of interest, we fit hierarchical linear models with a full random effects structure, which included varying intercepts and – slopes and full variance-covariance matrices for the random effects (Barr, Levy, Scheepers, & Tily, 2013).

The model specifications used in all analyses were as follows. A hierarchical linear model assumes that each data point (i.e. the mean ERP amplitude in the time window of interest) a_i is generated from a normal distribution with mean μ_i and standard deviation σ :

$$a_i \sim N(\mu_i, \sigma)$$

where μ_i is defined as a hierarchical linear regression:

$$\mu_i = \beta^T \mathbf{x}_i + \beta_{\text{subj}[i]}^T \mathbf{x}_i + \beta_{\text{item}[i]}^T \mathbf{x}_i$$

where \mathbf{x}_i is a vector containing the contrast coding of the fixed effects (including the intercept), β is a vector containing the estimates of the respective fixed effects, and $\beta_{\text{subj}[i]}$ and $\beta_{\text{item}[i]}$ contain the random effects estimates of the subject or item that produced the i^{th} data point.

This assumption about the generative process underlying the data is referred to as likelihood. We estimated the parameters of this likelihood from the data: the fixed effect estimates β , the random effects estimates β_{subj} and β_{item} and the standard deviation σ . The fixed effect estimates β were the parameters of theoretical interest for our research questions. For each parameter we assumed some prior distribution which specified the a priori probability of different parameter values. We used weakly informative priors which have a regularising effect on the parameter estimates, meaning that more conservative effect estimates are preferred (for discussion, see Gelman et al., 2014). Specifically, as a prior for the fixed effect intercept parameter β_0 we used a normal distribution with a mean of zero and a standard deviation of ten $N(0, 10)$. The value of 10 for the standard deviation was intended to allow for a wide range of possible estimates, as the N400 overall magnitude varies largely across studies depending on multiple experimental factors (e.g. SOA, presentation modality). Given this standard deviation, 95% of the probability mass of the prior covered the interval ranging from -20 to 20μ V. As priors for all other fixed effects we used a standard normal distribution $N(0, 1)$. The standard deviation of the fixed effects was smaller than that of the intercept, such that 95% of the probability mass of the prior covered a range from -2 to 2μ V. This was the range of semantic priming effect sizes that we deemed likely given previous studies. As a prior for sigma we used a standard normal distribution truncated at zero $N_+(0, 1)$, in order to ensure positive values for sigma.

As a prior for the random effects we used a multivariate normal distribution $MVN(0, \text{Cov})$ with a mean of 0 and a covariance matrix Cov , where Cov can be expressed in terms of a diagonal matrix Var and a correlation matrix R containing the variances and correlations of the random effects parameters: $\text{Cov} = \text{Var} \times R \times \text{Var}$ (for details, see McElreath, 2015). As priors for the variance of each random effect we used a truncated standard normal distribution $N_+(0, 1)$ and as a prior for the correlation matrix R we used a so-called LKJ prior, i.e. a distribution over possible correlation matrices (Joe, 2006; Lewandowski, Kurowicka, & Joe, 2009). The LKJ distribution has one parameter, η , which controls how much probability mass is assigned to extreme correlations (e.g. setting η to 1 means that

all correlation matrices are equally likely; for more detail, see Stan Development Team, 2017b). In all our analyses, we set η to 2 in order to slightly disfavour extreme correlations, following recent recommendations for doing Bayesian analyses in psycholinguistic experiments (Sorensen, Hohenstein, & Vasishth, 2016).

To address the research question, we fit two models with different fixed effect structures. First, we fit a model with a main effect of *possessive type* (NP/pronoun) and two pairwise comparisons that estimated semantic priming (or relatedness) effects for the possessive NP and pronoun conditions separately. The conditions with possessive NPs before the target words (1a, b) were coded as 0.5 and the conditions with possessive pronouns before the target words (1c,d) were coded as -0.5. In this and all following analyses, the unrelated conditions were coded as 0.5 and the related conditions as -0.5, with facilitatory priming always being indicated by negative effects. Note that the main effect of possessive type was not of theoretical interest, as we did not have any predictions as to whether N400 responses to target words overall (collapsing across related and unrelated conditions) would vary depending on whether they were preceded by a NP or a pronoun, which vary on several lexical variables such as length, frequency and imaginability. Rather, this effect was included to account for the variance in the data associated with this factor.

The research questions addressed by the first model were whether there was a priming effect for NPs and whether there was a priming effect for pronouns. Second, to address the question of whether the pronoun related condition differed from the coreference control condition, we fit a model that directly compared the related pronoun condition, coded as -0.5, with the coreference control condition, coded as 0.5.

To facilitate comparison with previous work that used by-subject ANOVAs and a null-hypothesis testing approach, all experiments were also reanalyzed using Type III SS repeated-measures ANOVA. These results are presented in the *Supplementary Materials*.

Results

Mean accuracy in the comprehension questions was 91% (SD = 5.1%). For the Bayesian analyses, we calculated the marginalised posterior distribution of each fixed effect, reflecting the probability of different effect sizes given the statistical model and the observed data. We present the mean of these posterior distributions and 95% credible intervals (CrI), which represent the interval where it is 95% certain that the true effect lies given the data. The analysis is considered to provide strong evidence for the presence of an effect

when the 95% CrI of the effect of interest does not include 0. Finally, we present the posterior probability of there being a facilitatory priming effect, which for the relatedness factor represents the posterior probability of a less negative N400 response for related than unrelated target words. In the remainder of the manuscript, the shorter denomination *probability of a priming effect* is used for conciseness.

ERP responses to target words are shown in Figure 1. The statistical analysis revealed some indication of priming for both the NP and pronoun conditions. The mean of the posterior distribution of the priming effect within the NP conditions was $-0.39 \mu\text{V}$ with a credible interval of $[-1.22, 0.47]$ and a probability of a priming effect of 0.813. The priming effect in pronouns was very similar: The mean of the priming effect's posterior distribution was $-0.38 \mu\text{V}$ with a credible interval of $[-1.19, 0.47]$ and a probability of a priming effect of 0.810. Finally, we note that N400 responses after possessive NPs were overall more positive than after possessive pronouns: The mean of the posterior distribution of the possessive type effect was $0.61 \mu\text{V}$ with a credible interval of $[-0.05, 1.27]$ and a probability of a negative effect of 0.037.

Crucially, the second model showed some evidence of a difference between related pronouns and the coreference control condition, as evidenced by a more negative N400 response in the coreference control condition. The mean of the posterior distribution was $-0.72 \mu\text{V}$ with a credible interval of $[-1.54, 0.12]$ and a probability of a negative effect of 0.954.

Discussion

The results of Experiment 1 were consistent with the experimental predictions. In the repeated NP conditions, words semantically related to the preceding NP showed facilitated N400 responses as compared with unrelated words. Moreover, quantitatively similar priming effects were obtained with coreferential pronouns. This supports the claim that upon encountering a coreferential pronoun such as *his*, readers rapidly reactivate the lexico-semantic features of its antecedent (e.g. *the prince*), which in turn preactivate semantically related words and thus reduce processing cost when one of these words is later encountered. Critically, the priming effect with coreferential pronouns was unlikely to result from residual activation of the NP *the prince* in the sentence context, as the comparison between the related pronoun and coreference control conditions showed that semantically related target words no longer displayed facilitated N400 responses when the pronoun referred to the proper

Experiment 1

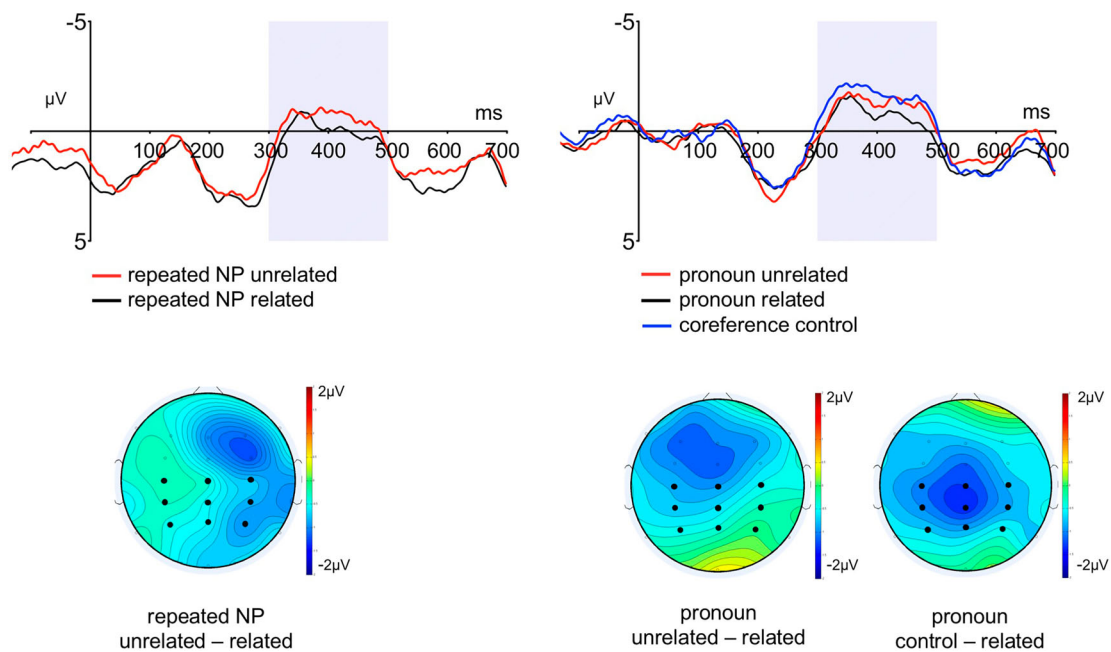


Figure 1. Experiment 1 (sentence paradigm): mean N400 responses to target words in the 300–500 ms time window (shaded) averaged across the 9 centro-parietal electrodes Cz, CPz, Pz, C3, CP3, P3, C4, CP4, P4 (bolded).

name antecedent. Therefore, as the context sentence was identical between these two conditions, priming effects were more likely attributable to coreference than to general context-priming effects.

However, although these results were consistent with the experimental predictions, they did not yield strong evidence of priming. This is clear in the credible intervals for both repeated NPs and pronouns, which included 0. Particularly surprising was the small magnitude of the priming effect in the repeated NP conditions, which was designed to provide a baseline measure of semantic relatedness. In order to account for these relatively small priming effects, we considered two explanations. The first was that semantic priming effects may be indeed very small in sentence contexts, as suggested by previous work in English (Camblin et al., 2007; Morris, 1994; Traxler et al., 2000). Although most of this previous work was done with eye-tracking, if small priming effects also affect ERP measures, this might have reduced experimental power in Experiment 1, limiting our ability to detect priming effects. Alternatively, it is possible that the semantic relationship between the critical word pairs (e.g. *prince–castle*) was not strong enough to differentially modulate participants' brain responses, despite having produced clear differences in their behavioural ratings (see *Materials*). The following studies addressed these possibilities.

Experiment 2

Experiments 2 and 3 were designed to address one of the possible explanations for the small priming effects in Experiment 1, namely that this experiment did not have enough trials to detect priming effects. With this goal, we divided the five experimental conditions into two separate experiments, allowing us to increase the number of trials per condition. We also modified the presentation of the preamble sentence and increased the difficulty of the comprehension questions to encourage careful reading. Experiment 2 focused on addressing the reliability of priming effects within the repeated NP conditions.

Methods

Participants

Thirty-two volunteers participated, with no exclusions, and were entered in the analysis (mean age = 20 years, age range = 18–29 years, 18 females).

Materials

Experiments 2 and 3 used a subset of 90 passages from the 150 in Experiment 1, focusing on those whose NP antecedent was strongly gender-biased and markedly related to the related target word. The two conditions with repeated NP possessives were tested in Experiment 2, and the three pronoun conditions were tested in Experiment 3. This

resulted in 45 trials per condition per participant in Experiment 2, and 30 trials per condition per participant in Experiment 3. In both experiments, no item was shown in more than one condition to the same participant. Only minor changes were made to the item sets, in order to improve the felicity of the discourse scenario. We also balanced the plurality of the target word after discovering that target words in Experiment 1 were more often plural in unrelated conditions than in related conditions. Further, the number of filler items was reduced to 60 and revised to be more engaging. Lastly, the difficulty of the comprehension questions was increased to encourage depth of processing (Love & McKoon, 2011; Sanford & Sturt, 2002; Stewart, Holler, & Kidd, 2007). As in Experiment 1, experimental and filler items were combined and randomised.

Procedure and analysis

Experiment 2 followed the same procedure and analysis as Experiment 1 with a few exceptions. First, the entire preamble sentence appeared on screen and participants read it at their own pace, pressing a button to initiate the RSVP presentation of the second sentence. Informal beta testing revealed that participants found this method of presentation more pleasant and less tiring than full RSVP of the two-sentence text. Second, participants were given feedback on their comprehension question accuracy at the end of each block to encourage attention

throughout the task. Materials were divided into five blocks that lasted approximately 6–8 min each, with in-between breaks. An entire experimental session lasted approximately 40 min. The equipment was identical to Experiment 1, except that new caps from Electro-Cap International were used for this and all following ERP experiments. The electrode configuration for these caps used a second frontal electrode (FP2) instead of the central occipital electrode (Oz). Approximately 8.8% of the trials were rejected due to excessive artifacts (range 0.3%–36.1%). Channel interpolation did not affect any of the electrodes of interest.

The analysis was similar to Experiment 1, with ERP responses being averaged across the same nine electrodes 300–500 ms after the presentation of the target word. Bayesian hierarchical linear models were fit with the same random effects structure and the same priors as in Experiment 1. In order to address the research question of whether there was a priming effect with possessive NPs, the model included a fixed effect of priming within the NP conditions.

Results

Mean accuracy for the comprehension questions was 93.9% (SD = 5.1%). ERP responses to target words are shown in Figure 2 (left panel). As in Experiment 1,

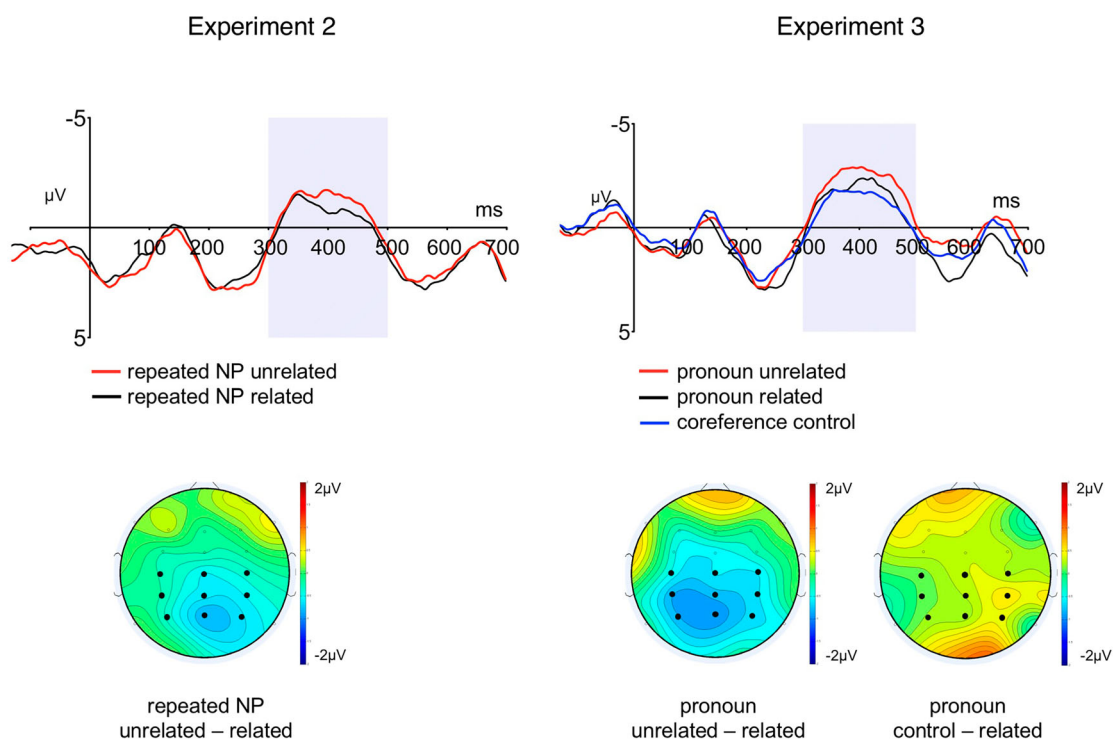


Figure 2. Experiments 1 and 2 (sentence paradigm): mean N400 responses to target words in the 300–500 ms time window (shaded) averaged across the 9 centro-parietal electrodes Cz, CPz, Pz, C3, CP3, P3, C4, CP4, P4 (bolded). The left panel shows the results of Experiment 2, and the right panel shows the results of Experiment 3.

there was only some indication of a priming effect: The mean of the priming effect's posterior distribution was $-0.35 \mu\text{V}$ with a credible interval of $[-1.01, 0.35]$ and a probability of a priming effect of 0.855.

Discussion

The results of Experiment 2 were consistent with those of Experiment 1: target nouns after semantically related possessive NPs showed reduced N400 amplitudes relative to semantically unrelated words. However, although Experiment 2 attempted to increase the number of trials by focusing on the possessive NP conditions, the statistical results showed only some indication of priming. In fact, the mean effect size estimates of priming in Experiments 1 and 2 were remarkably similar ($-0.35 \mu\text{V}$ and $-0.39 \mu\text{V}$, respectively). This suggests that small priming effects with possessive NPs may result indeed from a reduced influence of semantic priming in sentence contexts, rather than an insufficient number of trials in Experiment 1. Experiment 3 addressed whether this conclusion extended to priming effects with coreferential pronouns, by focusing only on conditions (1c,d,e) of Experiment 1.

Experiment 3

Methods

Participants

Fifty-seven volunteers participated in Experiment 3, but nine were excluded due to a technical error in the experiment presentation. Of the remaining 48 participants, six were excluded due to excessive artifacts, such that forty-two participants were included in the analysis (mean age = 21 years, age range = 18–27 years, 24 females).

Materials, procedure and analysis

Experiment 3 used the same 90 item sets and procedure as Experiment 2, but it only presented the three pronoun conditions. Approximately 8.9% of trials were rejected due to excessive artifacts (range 0%–34.4%). Channel interpolation did not affect any of the electrodes of interest. The item sets were distributed across three lists in a Latin Square design, combined with the sixty filler items and randomised. Bayesian analyses were performed using maximal random-effects structures and the same weakly-informative priors as in previous experiments. The first statistical model included a fixed effect of priming and the second model directly compared the related pronoun and coreference control conditions. As in Experiment 1, the related pronoun condition was coded as -0.5 and the coreference control condition as 0.5 , such that a

negative effect indicates more facilitation for the related pronoun than the coreference control condition.

Results

Main analyses

Mean accuracy in the comprehension questions was 90.1% ($SD = 4.9\%$). ERP responses to target words are shown in Figure 2 (right panel). The results in the coreferential pronoun conditions were consistent with Experiment 1, and it revealed only some indication of priming, with the mean of the posterior distribution of the priming effect being $-0.51 \mu\text{V}$ with a credible interval of $[-1.27, 0.25]$ and a probability of a priming effect of 0.901.

The results of the second model, however, differed from that of Experiment 1. The analysis revealed no evidence of a difference between the coreference control and the related pronoun condition. The mean of the effect's posterior distribution was positive with a value of 0.31, a credible interval of $[-0.47, 1.05]$ and a posterior probability of a negative effect of 0.218.

Supplementary analyses

The results of Experiment 3 differed from those of Experiment 1 in the behaviour of the coreference control condition. In Experiment 1, the pronoun related condition showed priming when compared to the coreference control condition, but in Experiment 3 there was no evidence of such difference, with the numeric pattern being the opposite. One explanation for the variable behaviour of the coreference control condition is that the gender marking of the pronoun did not render it fully referentially unambiguous, such that even in this condition, readers may have still reactivated the occupation antecedent for some items, yielding semantic priming and thus reducing or obscuring the N400 contrast between the pronoun related and the coreference control conditions. In other words, it is possible that in the coreference control condition of some items, such as "*Nicole stopped the president in the hallway before the broadcast. She feared that her administration ...*", readers still reactivated (at least partially or temporarily) the antecedent "*the president*", despite the mismatch between the feminine gender of the pronoun and the male bias of the occupation antecedent.

Note that although the gender bias of the occupation antecedents was verified in a separate norming study (see Experiment 1), we did not measure its downstream consequences in our sentence contexts. To address this shortcoming, we conducted an additional rating study to quantify the referential ambiguity of the coreference control condition, and we used the ratings as a by-item predictor of the ERP responses to see if

they could explain some of the variability in the N400 differences between the pronoun related and coreference control conditions in Experiments 1 and 3. The format of the rating experiment followed Kehler, Kertz, Rohde, and Elman (2008: Experiment 1). 150 participants (mean age = 32 years, age range = 19–56 years, 67 females) read sentence fragments in rapid serial visual presentation (SOA = 500 ms) until the target word, and then answered a forced-choice question targeting the referent of the pronoun: e.g. “Genevieve traveled with the prince for many weeks. She wished that *his* / *her* castle...”; “Who did the castle belong to? -Genevieve, -the prince”. Participants saw each item in either the pronoun related or coreference control conditions (e.g. “his castle” vs. “her castle”).

The results of the norming study revealed strong evidence of a contrast between the two conditions: the occupation antecedent was chosen as a referent on 92% of the trials in the pronoun related condition but only on 30% of the trials in the coreference control condition (mean of the posterior distribution = 4.67 log odds, CrI = [3.88, 5.65], probability of a negative effect = 1.00). However, there were differences across items in the referential ambiguity of the coreference control condition: whereas in some items the occupation antecedent was never selected, in other items it was selected on 50% or more of the trials. Thus, to assess whether this variability modulated the N400 differences between the coreference control and the pronoun related conditions, we entered the mean by-item ratings as an additional centered predictor to the model that targeted the contrast between these conditions (i.e. the second model presented in the *Main analyses* section). However, there was no statistical evidence of an interaction between the ratings and the difference between the coreference control and pronoun related conditions in Experiment 1 (mean of the posterior distribution = $-0.68 \mu\text{V}$, CrI = $[-2.38, 0.96]$, probability of a negative interaction = 0.771) or Experiment 3 (mean of the posterior distribution = $-0.19 \mu\text{V}$, CrI = $[-1.87, 1.50]$, probability of a negative interaction = 0.595). Therefore, we did not observe evidence that differential degrees of referential ambiguity in the control condition could explain the lack of a reliable difference between this condition and the pronoun related condition across experiments.

Discussion

The results of Experiment 3 only partially replicated those of Experiment 1. First, target words semantically related to the pronoun’s antecedent showed facilitated N400 responses compared to unrelated words. This priming effect replicated the one in Experiment 1 and it had a stronger mean amplitude ($-0.37 \mu\text{V}$ vs. $-0.50 \mu\text{V}$).

However, this result is unlikely to reflect facilitation due to antecedent reactivation, because the coreference control condition, which was intended to rule out a general effect of context-priming, did not show a larger N400 than the related pronoun condition: in fact, N400 responses were as much facilitated in the coreference control as in the related pronoun conditions. Therefore, it cannot be ruled out that the reduced N400 to words like *castle* resulted from residual activation from reading the NP *the prince* in the sentence context, rather than its retrieval specifically due to coreference. Before addressing the implications of this finding, we present two final experiments, which aimed to rule out a potential problem in the experimental materials.

Experiment 4

Experiments 1 and 2 showed surprisingly small N400 facilitation effects, even when possessive repeated NPs were used immediately before the target word. This observation limits the conclusions that can be drawn from the failure to find strong priming effects in the pronoun conditions, because it suggests that the “base” relatedness effect that the antecedent retrieval hypothesis would predict would have been quite small to begin with. However, it is not obvious why the N400 relatedness effect was so small in the NP conditions, given that the behavioural norming showed clear differences in relatedness. Experiment 4 was designed to shed more light on this puzzle. We conducted an ERP experiment using a “word pair” paradigm typical of classic semantic priming manipulations. Our goal was to determine whether semantically related words also showed small N400 priming effects in a word pair paradigm.

Methods

Participants

Thirty-two participants took part in the experiment, but three participants were excluded due to excessive artifacts and five were excluded due to poor accuracy in the memory test (see *Procedure*). Twenty-four participants were included in the final analysis (mean age = 20 years, age range = 18–27 years, 20 females).

Materials

Stimuli consisted of the 150 related (e.g. *prince–castle*) and unrelated (e.g. *prince – estate*) noun pairs from Experiment 1. The pairs were divided into two lists and intermixed with thirty fillers consisting of ten strongly related pairs (e.g. *migraine – headache*), ten unrelated pairs (e.g. *physician–flower*), and ten weakly related

pairs, (e.g. *helicopter – bicycle*). The order of experimental and filler items was randomised for each list.

Procedure and analysis

Word pairs were presented using the same stimulus-onset asynchrony interval (500 ms) as in the sentence experiments. To ensure continued attention during the experiment, participants performed a memory test after the experiment. The test consisted of ten pairs of nouns that had appeared together during the experiment and ten pairs of nouns that had appeared previously but not as a pair. Participants were told about the memory test prior to the experiment and its format was described to them. They were advised that although all the words in the test had been displayed previously, the critical question was whether they had been presented together as a pair. Participants who were less than 60% accurate in the memory test were excluded from analysis. The experimental session lasted 10–15 min, with two resting intervals. As the overall length of the experimental session was short, experiments 4 and 5 were preceded by another experimental session. The majority of the participants in each group participated in an unrelated word pair experiment (75% of participants in experiment 4, 85% in experiment 5), and the remaining participants took part in an unrelated sentence experiment. Data was recorded

and preprocessed as in the preceding experiments. Approximately 15.5% of the trials were rejected due to artifacts (range 0%–40%). Channel interpolation did not affect any of the electrodes of interest.

We performed a similar Bayesian analysis as for the preceding experiments. As before, we aggregated the data over the nine electrodes of interest and fit a hierarchical linear model with a fixed effect of priming, a full random effects structure and the same weakly informative priors as in previous analyses.

Results

Mean accuracy in the memory test was 75.9% (SD = 7.7%). ERP responses to target words are shown in Figure 3 (left panel). The analysis of the ERP responses revealed strong evidence of priming in the N400 time-window, and numerically, the effect appeared to onset even earlier.¹ The mean of the priming effect's posterior distribution was $-1.36 \mu\text{V}$ with a credible interval of $[-1.95, -0.77]$ and a probability of a priming effect of 1.000.

Discussion

Experiment 4 showed a large N400 priming effect for the pairs that served as the antecedent and target word in

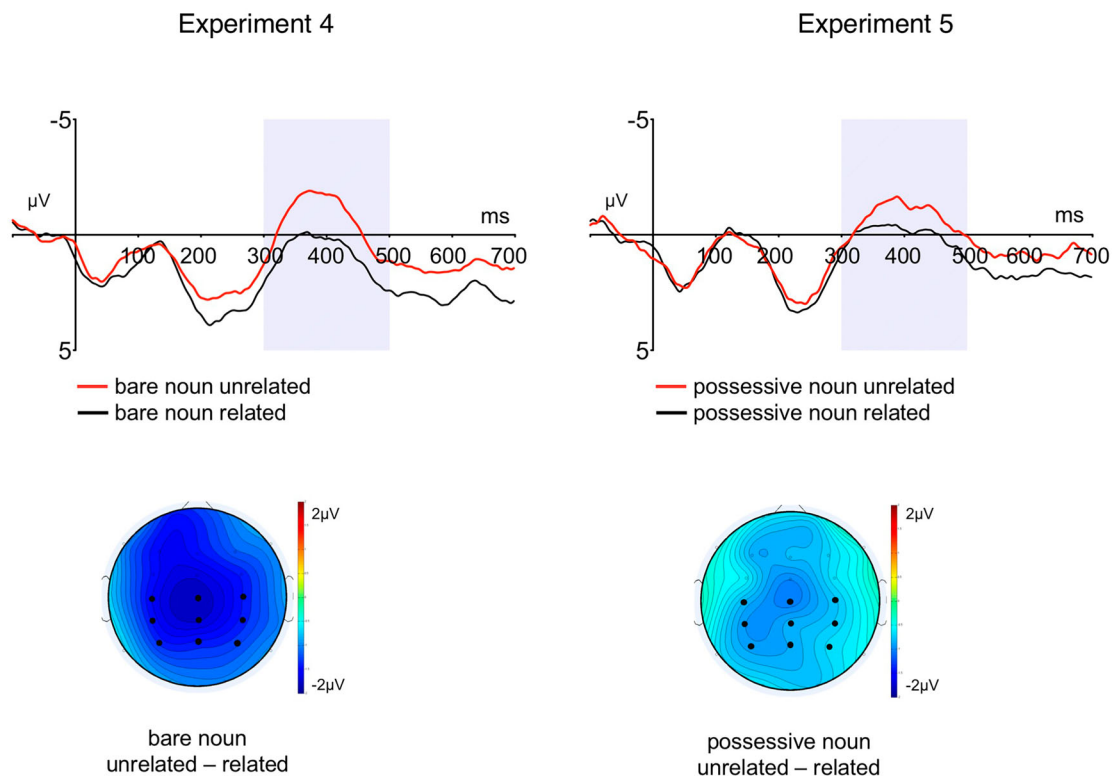


Figure 3. Experiments 4 and 5 (word pair paradigm): mean N400 responses to target words in the 300–500 ms time window (shaded) averaged across the 9 centro-parietal electrodes Cz, CPz, Pz, C3, CP3, P3, C4, CP4, P4 (bolded). The left panel shows the results of Experiment 4, and the right panel shows the results of Experiment 5.

the sentence experiments. These results replicate prior work that has reported a reduction in N400 amplitude for words following semantically related primes (Brown et al., 2000; Holcomb, 1988; Kutas & Van Petten, 1994; Rugg, 1985). Notably, the priming effect was estimated to be larger than that observed for exactly the same pairs in Experiments 1 and 2, which used the same stimulus-onset asynchrony between the “prime” and “target” words.

There are at least two explanations for this difference across experiments. First, prior literature has sometimes reported that N400 effects of semantic relatedness are smaller in sentence contexts than in word pair paradigms (Coulson, Federmeier, Van Petten, & Kutas, 2005; Ledoux, Camblin, Swaab & Gordon, 2006). Therefore, the smaller effect in Experiments 1 and 2 may be due to the fact these studies displayed the critical words within sentence contexts. This would suggest that coreference paradigms involving sentences inevitably yield small N400 relatedness effects. But a second possibility relates to the fact that prime words in the sentence experiments were marked with possessive morphology (e.g. *prince's*), in contrast to the word pair experiment, in which prime words appeared without case morphology (e.g. *prince*). Specifically, it is possible that the possessive marker on the prime word altered its effect on the target word. This could occur if possessively-marked nouns are costlier to process and thus delay or interfere with semantic priming, or if the possessive marker on the prime word alters the pre-activation process, resulting in different words being preactivated due to the occurrence of possessive vs. bare case primes. To address this possibility, Experiment 5 used a word pair paradigm identical to Experiment 4 but added possessive morphology to the prime word.

Experiment 5

Experiment 5 was designed to examine the presence of N400 facilitation effects in word pairs where the prime carried possessive morphology. Therefore, the same word pair paradigm of Experiment 4 was used, but the genitive marker's was added to the prime word.

Methods

Participants

Twenty-nine volunteers participated, but nine participants were excluded from analysis, two due to excessive artifacts and seven due to accuracy below 60% in the behavioural task. Twenty participants were included in

the final analysis (mean age = 20 years, age range = 18–23 years, 15 females).

Materials, procedure and analysis

Materials were identical to those presented in Experiment 4 with the addition of the possessive marker to the prime words. Approximately 13.43% of the trials were rejected due to artifacts (range: 2.67%–38.67%). Channel interpolation only affected one channel of interest: (participant “5–17”, channel C3). The same Bayesian model as in Experiment 4 was fit to the data.

Results

Mean accuracy in the memory test was 73.0% (SD = 9.8%). ERP responses to target words are shown in Figure 3 (right panel). The analysis of ERP responses revealed evidence of priming: the mean of the fixed effect's posterior distribution was $-0.76 \mu\text{V}$ with a credible interval of $[-1.43, -0.01]$ and a probability of a priming effect of 0.977.

Discussion

Experiment 5 examined whether possessively-marked primes facilitated N400 responses for semantically related following words. The results supported this hypothesis, but they also showed that the mean size of the N400 relatedness effect was reduced compared to the bare primes used in Experiment 4 ($-0.76 \mu\text{V}$ vs. $-1.37 \mu\text{V}$, respectively). However, a potential caveat about the comparison between Experiments 4 and 5 is the fact that the numerical divergence between the related and unrelated conditions was notably earlier in Experiment 4. One might be concerned that this earlier divergence reflects a distinct effect that overlaps with the N400 time-window, creating the impression of a larger N400 effect size. Although we cannot rule out this possibility, we believe it is unlikely, both because large N400 effects often onset prior to 300 ms in the literature (e.g. Federmeier & Kutas, 1999; Holcomb, 1988; Lau, Holcomb, & Kuperberg, 2013), and because the centro-parietal distribution of the effects in both the 200–300 ms window and the 300–500 ms window are similar to those standardly reported for N400 effects.

A second potential caveat about the comparison between Experiments 4 and 5 is that it was based only on the numeric estimates of the priming effects. Therefore, in order to provide a more direct comparison and to precisely estimate the size of the N400 difference between Experiments 4 and 5, we conducted an analysis of the data pooled across experiments, as described below.

Pooled analysis

To address the variability of the findings across experiments, we conducted an analysis of their pooled data. Our goal was to achieve higher statistical power and hence to get more precise estimates for the experimental effects. Our questions of interest were: (1) whether the five pooled experiments offered strong evidence of a priming effect for full NPs; (2) whether there was strong evidence of priming for coreferential pronouns; and (3) whether the coreference control condition reliably differed from the pronoun related condition, consistent with the priming effect for pronouns being specifically due to coreference, rather than a more general context priming effect.

The pooled analysis also allowed us to address two additional observations that emerged during the experiments about the role of the experimental paradigms and of the case morphology of the prime words. Specifically, we wanted to examine: (4) whether priming effects with possessive NP primes were reliably smaller within sentences than word pair paradigms, and (5) whether priming effects with NP primes were stronger with bare than with genitive primes. Pooling the data allowed us to directly assess these questions by examining whether there were interactions between priming and paradigm, on the one hand, and between priming and case, on the other.

In order to address the five questions outlined above, the following models were coded. The first model included as fixed effects a priming effect within full NPs (relevant for question 1), a priming effect within pronouns (relevant for question 2), an interaction between paradigm and priming (relevant for question 4), and an interaction between case and priming (relevant for question 5). Note that the two interactions were only computed over trials that contained NP primes, as only these were presented in both sentence and word pair paradigms and manipulated for case. Furthermore, the interaction between paradigm and priming was only computed over the possessive trials, since having bare case primes in sentences would have rendered them ungrammatical (e.g. *She wished that the *prince castle wasn't so far away*). In addition, three other fixed effects were used: a main effect of possessive type (NP/pronoun), a main effect of paradigm (sentence/word pair) and a main effect of case (bare/possessive). These effects were not of theoretical interest because they were orthogonal to the research questions. For example, the main effect of paradigm quantified whether trials with bare case primes elicited more negative N400 responses than trials with possessive primes (regardless of the factor relatedness). Since our research

question was whether case morphology enhanced (or diminished) priming effects specifically, overall effects of case (or paradigm or possessive type) were not of interest. However, these effects were included to reflect the experimental design, and thus to account for the variance associated with these manipulations and because the interpretation of the critical interactions would not be possible otherwise.

Fixed and random effects were coded consistently with the individual experiments. For the factor *priming* (or *relatedness*), related trials were coded as -0.5 , and unrelated trials as 0.5 , such that a negative estimate reflected facilitatory priming. For the factor *possessive type*, trials with coreferential pronouns as primes were coded as -0.5 and trials with NP primes as 0.5 . For the factor *paradigm*, sentence trials were coded as -0.5 and word pair trials as 0.5 . For the factor *case*, trials with possessive primes were coded as -0.5 and trials with bare case primes were coded as 0.5 . The random-effects structure of the model included the effects of priming (for both NPs and pronouns) and the effect of possessive type as within-subject comparisons. All other fixed effects were between-subject comparisons, as they had been manipulated across subjects. The pooled analyses used the same weakly informative priors as the individual experiments.

In order to follow-up on the interactions between priming and case and priming and paradigm, two additional models were fit, which replaced the interactions with the corresponding pairwise comparisons. Lastly, to address question (3), regarding the difference between the related pronoun and the coreference control conditions, the data was subset to include only these conditions and their difference was coded as a fixed effect together with a full random effects structure, with the same weakly informative priors as in the previous models.

Results

The results of the pooled analysis are summarised in [Table 1](#) and [Figure 4](#). Note that only the effects of theoretical interest are discussed in the text, although the table presents all estimates for completeness. As with the individual analyses the term *probability of a priming effect* is used as a shorthand for the posterior probability of there being a facilitatory priming effect, representing the probability of a less negative N400 response for related than unrelated target words.

The pooled analysis showed strong evidence of priming when target words were preceded by related NP primes. There was also some indication of priming

Table 1. Statistical results from the pooled analysis. Each effect is presented together with the mean of its posterior distribution (μV), a 95% credible interval (μV), and the probability of there being a facilitatory effect. Effects whose credible interval excludes zero are bolded, and they are described in the text as providing strong evidence for the effect of interest.

Effect	Posterior mean	Credible interval	Prob(effect < 0)
Priming: NPs	-0.91	[-1.27, -0.54]	1.000
Priming: coreferential pronouns	-0.50	[-1.06, 0.07]	0.960
Coreference control	-0.15	[-0.77, 0.47]	0.686
Priming \times Paradigm: NPs	-0.45	[-0.92, 0.02]	0.949
Priming: sentences	-0.42	[-0.97, 0.12]	0.938
Priming: word pairs	-1.81	[-2.70, -0.90]	1.000
Priming \times Case: NPs	-0.50	[-1.00, -0.02]	0.980
Priming: bare case primes	-1.35	[-1.92, -0.78]	1.000
Priming: possessive case primes	-1.06	[-1.89, -0.21]	0.991
Possessive type	0.59	[-0.08, 1.25]	0.045
Paradigm	0.54	[-0.23, 1.37]	0.087
Case	0.46	[-0.48, 1.41]	0.170

in the coreferential pronoun conditions, although this priming effect was smaller than for the NP conditions, with a broader posterior distribution and a 95% credible interval that included 0. Furthermore, the model comparing the coreference control condition with the related pronoun condition did not support a reliable difference between conditions. Taken together, these results do not provide strong evidence that semantic facilitation with pronouns are driven specifically by coreference.

Interestingly, there was some evidence for an interaction between priming and paradigm. Pairwise comparisons showed that this interaction was due to a larger priming effect when the target stimuli were presented as word pairs as compared to within sentences. Furthermore, there was strong evidence of an interaction between case and priming, confirming the numeric contrast between Experiments 4 and 5: Priming effects were reduced when the prime word was presented with possessive case, compared with a bare case presentation. Overall, these results show that semantic priming effects are reduced by both sentence context and possessive case morphology.

General discussion

This study examined the mechanisms supporting pronoun interpretation during coreference. We asked whether the retrieval of the antecedent of a pronoun could be indexed by semantic priming effects, as suggested by previous cross-modal studies (Leiman, 1982; Nicol, 1988; Shillcock, 1982). According to these studies, if readers rapidly reactivate the lexical representation of an antecedent upon encountering a pronoun, then words that are semantically associated

to the antecedent should become preactivated, resulting in processing facilitation when one of these words is later encountered. However, cross-modal priming measures rely on an explicit decision task and are thus potentially prone to task-related strategies. In the current study, the N400 effect was used to provide an implicit time-sensitive diagnostic of (the consequences of) antecedent retrieval, in the absence of an overt decision component.

Overall, our results did not provide reliable evidence of semantic facilitation specifically due to coreference. First, the analysis of the combined experiments only yielded weak evidence of N400 reductions for words semantically related to the pronoun's antecedent (e.g. *his castle* vs. *estate*). Second, N400 responses in the semantically related condition (e.g. *his castle*) did not reliably differ from the condition in which the pronoun no longer referred to the target antecedent (*her castle*, coreference control condition). The coreference control condition was designed to ensure that readers' facilitated processing of the target words was specifically due to coreference, rather than from having read the semantically associated noun *prince* in the prior context. Therefore, the absence of a clear contrast between the pronoun related and the coreference control conditions suggests that the weak facilitation effects for the former may have been more due to general context-priming effects than to antecedent reactivation. In what follows, we discuss the implications of our findings for psycholinguistic accounts of pronoun processing, focusing on the role of antecedent retrieval during coreference. Then, we turn to some practical limitations of the semantic facilitation paradigm as a tool to study coreference, as well as the additional findings that it yielded about the influence of context and case morphology on semantic priming effects.

What is accessed during pronoun processing?

Our study was motivated by previous cross-modal work, which had reported semantic priming effects immediately after the presentation of a pronoun. However, a longstanding concern about the cross-modal paradigm is that it is prone to strategic effects, as detecting semantic relationships between pronouns and their antecedents can help participants improve their task performance. The current results are consistent with our failure in prior work to observe antecedent relatedness effects in eye-tracking-while-reading in English (Lago et al., 2017). Similarly, another recent ERP study failed to show effects of antecedent concreteness at either the pronoun or the content words

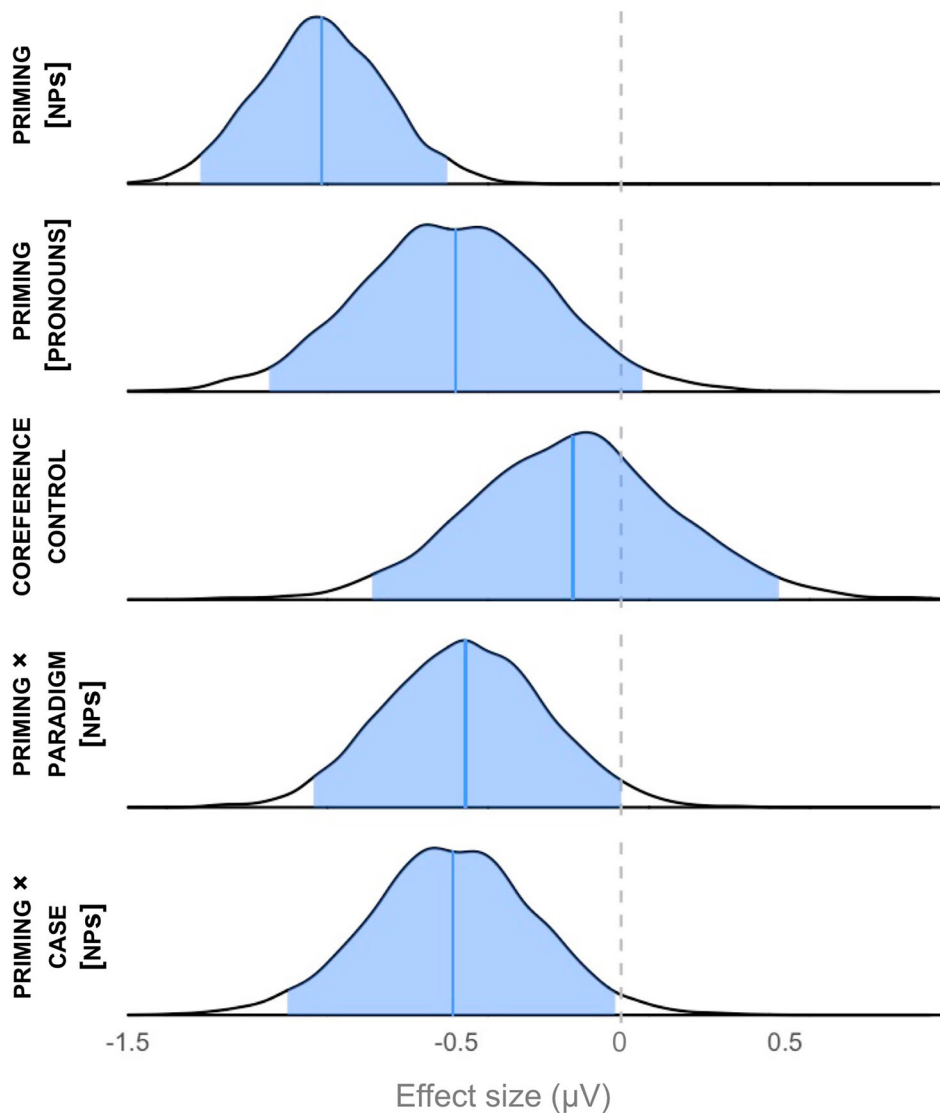


Figure 4. Posterior distributions of the effects of interest in the pooled analysis. The distributions display the posterior probability of different effect sizes, with vertical bars corresponding to posterior means and shaded areas to 95% credible intervals. The gray dashed line corresponds to an effect size of zero.

following it (Smith & Federmeier, 2015). Therefore, the fact that antecedent relatedness effects were not reliably observed in studies using implicit measures suggests that cross-modal priming effects in previous work may have partly resulted from task-related strategies, rather than the automatic retrieval of the pronoun's antecedent. Under this interpretation, two outstanding questions are why coreference in English might not elicit automatic semantic priming effects and what this entails for models of pronoun processing in comprehension.

One explanation for the lack of strong evidence of priming effects is that pronoun interpretation in English requires contact with the referent of the pronoun in speakers' discourse model but does not

involve the retrieval of the antecedent's representation from long-term memory. Specifically, a previous eye-tracking study by Lago et al. (2017) showed evidence of semantic facilitation effects in German (a language in which pronouns carry grammatical gender) but not in English. To account for the cross-linguistic contrast, it was proposed that the retrieval of discourse referents and linguistic antecedents during coreference was conditioned by the grammatical properties of each language. In languages without grammatical gender like English, the features necessary to identify an antecedent, which include conceptual gender, would all be bound to the discourse representation of the pronoun's referent (Cloitre & Bever, 1988; Lucas et al., 1990), thus obviating the need for long-term memory retrieval and

preventing the occurrence of spreading activation between semantically-related lexical items. This is because the process of spreading activation is assumed to rely on differentially weighted connections between long-term memory units, and thus it may not happen without these units' reactivation (Collins & Loftus, 1975; Forster, 1976; Levelt et al., 1999; Morton, 1979). By contrast, in languages with grammatical gender, like German, the syntactic gender of an antecedent often has no conceptual correlates and might be solely stored in the lexicon. In this case, the reactivation of the antecedent representation in long-term lexical memory would also be necessary to identify an appropriate antecedent and to license the agreement features of the pronoun.

Under this account, the lack of reliable semantic priming effects occurred because the English pronouns of the current study did not reactivate antecedent representations in comprehenders' long-term memory. Rather, identifying an appropriate referent for an English pronoun could be achieved by probing the most salient discourse referents for their conceptual gender. Critically, discourse referents are either short-term memory representations or "recently constructed" long-term memory representations (e.g. *the prince* in the current discourse may have idiosyncratic properties distinct from the more general concept of *prince*). Therefore, there is no reason to think that discourse referents would have the long-term connections weighted by past experience that would yield spreading activation effects (e.g. Smith & Federmeier, 2015).

This idea is consistent with prior work suggesting that even full noun phrase references to a discourse entity do not continue to activate the features associated with the lexical item that introduced them initially. Most famously, when the prior discourse establishes that a peanut is romantically entangled, further references to *the peanut* drive rapid facilitation for predicates like "in love" relative to lexically associated predicates like "salted" (Nieuwland & Van Berkum, 2006). Therefore, if nouns themselves often do not elicit reactivation of the full set of lexical features associated with an antecedent noun in long-term memory, it is unclear why pronouns should do so, unless as a byproduct of a grammatical requirement (as proposed for languages with grammatical gender).

The current results leave open interesting questions about the nature of the memory operations needed to link a pronoun with a discourse referent. These operations depend on the type of memory architecture involved in language processing. Early accounts proposed a multistore system, which distinguished between long- and short-term memory stores (Baddeley,

1986, 1992, 2000; James, 1890; Repov & Baddeley, 2006). Within this type of multistore architecture, current discourse referents might be represented in short term memory, and coreference in English would involve selecting among a set of referents on the basis of morpho-syntactic information, world knowledge, and discourse structure (e.g. Garrod & Sanford, 1982).

Alternatively, more recent memory accounts do away with the distinction between long- and short-term stores. Instead, these accounts propose that all items must be retrieved from a single memory store, with the exception of a very limited set of items that are currently under processing, which are assumed to be held in speakers' focus of attention (Anderson et al., 2004; Cowan, 1988, 2000; McElree, 2001; Oberauer, 2002). In these architectures, one could capture the presence or absence of semantic facilitation effects in several ways. One way would propose that certain classes of memory representations are organised into associative networks that yield spreading activation (i.e. words) whereas others are not (i.e. the conceptual features bound to discourse referents). Whereas this proposal is logically possible, we are not aware of any independent functional or neurobiological motivations for it. Alternatively, it could be proposed that retrieving the memory representation of a discourse referent does not automatically trigger retrieval of the conceptual features that are bound to it. Rather, the retrieved object may be a discourse ID or label, such as "discourse referent 2", and the information bound to that referent (i.e. its conceptual features and predicated attributes) may be only retrieved when other parts of the sentence prompt an enriched interpretation. For example, the awareness of the unusual state of affairs described by the sentence *The farmer worked in a skyscraper* might not occur automatically, but rather because "working in a skyscraper" triggers a new interpretive goal of inferring the kind of work done by the farmer, which subsequently prompts reaccess of the conceptual features of the discourse referent that is the agent of the working event.

Finally, it is worth noting that an assumption that underlies the current work and much prior work is that antecedent access occurs immediately at the pronoun. However, this assumption can be challenged by findings that have shown pronoun interpretation to be sometimes temporally delayed or even skipped altogether (Carpenter & Just, 1977; Greene, McKoon, & Ratcliff, 1992; Love & McKoon, 2011; see also discussion by Sanford & Garrod, 1989). In addition, recent work has argued that pronoun interpretation can begin predictively, before a pronoun is even encountered. For example, comprehenders often have expectations about which discourse entity will be referred to next,

and these expectations can rapidly impact pronoun interpretation (Arnold, 2010; Rohde & Kehler, 2014). Although more work is needed to identify the variables that can affect the time course of coreference, these findings question the need for a retrieval process time-locked to the appearance of a pronoun, thus providing a possible explanation for the lack of immediate priming effects in the present study.

Semantic facilitation effects in sentence processing: beyond pronouns

Beyond the pronoun conditions that were of interest for the current study, we found that N400 facilitation effects were relatively small in sentence contexts, even when the semantic relationships occurred between two adjacent noun phrases (*the prince's castle vs. estate*). This resulted in smaller N400 priming effects in sentences (Experiments 1, 2 and 3) than in isolated word pair paradigms (Experiments 4 and 5). These results replicate previous eye-tracking and ERP studies, which have found smaller and/or delayed facilitation effects in sentences compared to word pairs (Camblin et al., 2007; Coulson et al., 2005; Carroll & Slowiaczek, 1986; Morris, 1994; Traxler et al., 2000; for review see Boudewyn, Gordon, Long, Polse, & Swaab, 2012). However, Experiment 5 showed that N400 reductions could also be partially attributed to the use of possessive morphology in sentence contexts, as genitive-marked primes clearly reduced the size of the N400 relatedness effects even in word pairs. Therefore, an important question is why possessive morphology and sentence context would reduce the size of priming effects.

With regard to the role of possessive morphology, one possibility is that the processing of the additional possessive marker was costlier and that this cost interfered with (or delayed) access to the lexical or conceptual features of the prime, in turn reducing the impact of word associations on the target. This hypothesis may seem to conflict with recent ERP findings suggesting that case information (e.g. accusative vs. nominative markings) takes a longer time to affect verb predictions than their lexico-semantic properties (e.g. Chow, Wang, Lau, & Phillips, 2018; Momma, Sakai, & Phillips, submitted). Nonetheless, case marking may play different roles (or affect the time-course of predictions differently) when implemented on nouns vs. pronouns (or when informing verb vs. noun predictions). Therefore, future work on the role of case marking in coreference is needed, for example by systematically manipulating the amount of time allocated to the presentation of a possessive pronoun, in order to give participants more time to process case morphology.

Alternatively, the possessive marker on the prime word may have affected how participants created expectations about following words, by generating predictions that did not necessarily consist of semantic associates. An advantage of this explanation is that it could potentially explain why the N400 effect may have been affected by both possessive morphology and sentence context. This explanation relies on classic theories of semantic priming, which proposed that priming effects in word pair paradigms had at least two different sources: automatic spreading activation and prediction (Neely, 1991; Posner & Snyder, 1975). Automatic spreading activation was proposed to drive much of the priming effect at shorter primer-target delays and/or when the prime was masked from consciousness. By contrast, more “controlled” processes like prediction would dominate at longer delays. Importantly, subsequent word-pair ERP studies demonstrated that a large part of N400 semantic facilitation effects at long delays was due to predictive processes, indicating that participants could rapidly (and perhaps implicitly) recognise the semantic associations between word pairs, and then use these relationships to predict which target would follow the prime (Brown et al., 2000; Lau et al., 2013).

Under this account, the use of bare-case primes in Experiment 4 may have forced participants to rely mainly on semantic association, which was the only source of information available to them. However, the use of possessive case, which entered the prime and target words into a structural relationship, as well as their embedding in sentence contexts may have made other sources of information more salient, reducing participants’ reliance on semantic association and thus the role of spreading activation mechanisms.

The broader implication of this explanation for the pronoun interpretation literature is that manipulations based on semantic association are unlikely to yield robust effects in sentence contexts, where predictions will be primarily dominated by structural and discourse information, more than by associations between individual lexical items. Therefore, studies aiming to use this approach in the future should exercise caution. It is also worth noting that in order to conclude that semantic facilitation effects on the target word reflect retrieval of information at the pronoun rather than prediction of the target itself, such designs must ensure that the contexts do not themselves predict the critical targets. On the other hand, if the question of interest is about the speed of antecedent identification rather than the types of representations involved, then two more promising paradigms are visual world eye-tracking, given its ability to detect rapid attentional changes, as well as ERP designs in which pronoun resolution shifts

predictions about upcoming words, given the broadly observed effects of prediction in N400 sentence studies (DeLong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004 but see Nieuwland et al., 2018).

Conclusion

Using an implicit and time-sensitive measure, the N400, we did not observe strong evidence of semantic facilitation effects specifically due to coreference in English. We believe that these data raise questions about the origins of such effects in cross-modal work and they raise the possibility that semantic relatedness effects in comprehension are not a cross-linguistically effective tool to probe for antecedent retrieval. Furthermore, the impact of linguistic information and sentence context on the size of N400 priming effects suggests that they may largely reflect the extent to which the upcoming input is predicted, rather than automatic spreading activation between lexical representations. We have also argued that prior work and broader theoretical considerations make it unlikely that English pronouns elicit the retrieval of the long-term memory representations that would trigger spreading activation to semantic associates of the antecedent. Rather, we suggest that pronoun interpretation involves either linking a pronoun with a discourse referent in working memory or retrieving a discourse referent without immediately reactivating its bound conceptual features.

Note

1. Although we had no prior hypotheses about priming effects prior to 300 ms, we conducted an exploratory analysis examining whether the mean ERP amplitude in the 200–300 ms time window differed between conditions. This analysis showed strong evidence that the unrelated condition was more negative than the related condition: mean of the posterior distribution = $-0.82 \mu\text{V}$, $\text{CrI} = [-1.38, -0.24]$, probability of a negative effect = 0.998. The topography of the effect was similar to that observed in the subsequent N400 time-window (see Supplementary Materials).

Acknowledgements

We thank Alexander Williams, Bruno Nicenboim and the audience of the CUNY Human Sentence Processing Conference in 2018 for comments and suggestions. We thank Laura Kertz, Hannah Rohde and Andrew Kehler for kindly sharing their experimental materials.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by a German Research Council project [grant number LA 3774/1-1] awarded to Sol Lago, and by a German Research Council project [grant number 317633480 – SFB 1287, Project Q] awarded to Shravan Vasishth and Ralf Engbert.

ORCID

Sol Lago  <http://orcid.org/0000-0002-4966-1913>

Anna Namyst  <http://orcid.org/0000-0001-6273-2205>

Lena A. Jäger  <http://orcid.org/0000-0001-9018-9713>

References

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, 111, 1036–1060.
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4), 187–203.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon.
- Baddeley, A. D. (1992). Working memory. *Science*, 255, 556–559.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods*, 39, 445–459.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Language and Cognitive Processes*, 27(5), 698–733.
- Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, 72, 158–190.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990.
- Bürkner, P.-C. (2016). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Cacciari, C., Carreiras, M., & Barbolini-Cionini, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, 37, 517–532. doi:10.1006/jmla.1997.2528
- Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, 56, 103–128. doi:10.1016/j.jml.2006.07.005
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic

- programming language. *Journal of Statistical Software*, 76 (1), 1–32.
- Carpenter, P. A., & Just, M. A. (1977). Reading comprehension as eyes see it. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 109–140). Hillsdale, NJ: Erlbaum.
- Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*, 14(6), 509–522. doi:10.3758/BF03202522
- Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5, 630. doi:10.3389/fpsyg.2014.00630
- Chow, W. Y., Wang, S., Lau, E., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition, & Neuroscience*, 33(7), 803–828.
- Cloitre, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes*, 3, 293–322. doi:10.1080/0169096880840209
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428. doi:10.1037/0033-295X.82.6.407
- Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 129–147.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163–191.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114.
- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effect on event-related potentials during reading. *Brain Research*, 1084, 89–103.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8 (8), 1117–1121.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, 8(3), 373–392. doi:10.1016/S0926-6410(99)00036-1
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146(1), 75–84. doi:10.1016/j.brainres.2006.06.101
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11.
- Frazier, L., Henstra, J., & Flores d' Arcais, G. B. (1996). Finding candidate antecedents: Phrase or conceptual entities. *University of Massachusetts Occasional Papers in Linguistics*, 19, 193–238.
- Garnham, A., Oakhill, J., Erlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language*, 34, 41–62. doi:10.1006/jmla.1995.1003
- Garrod, S. C., & Sanford, A. J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics*, 1, 21–41.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 266–283.
- Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, 35, 66–85.
- James, W. (1890). *Principles of Psychology*. New York: Henry Holt.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97, 2177–2189.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25, 1–44.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533–572.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (pp. 139–187). Greenwich, CT: JAI.
- Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic.
- Lago, S., Sloggett, S., Schlueter, Z., Chow, W. Y., Williams, A., Lau, E., & Phillips, C. (2017). Coreference and antecedent representation across languages. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 795–817. doi:10.1037/xlm0000343
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502.
- Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of intralexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews*, 5, 107–127.
- Leiman, J. M. (1982). *A chronometric analysis of referent assignment to pronouns* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38. doi:10.1017/S0140525X99001776
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended

- onion method. *Journal of Multivariate Analysis*, 100(9), 1989–2001.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213. doi:10.3389/fnhum.2014.00213
- Love, J., & McKoon, G. (2011). Rules of engagement: Incomplete and complete pronoun resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 874–887.
- Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory & Cognition*, 18, 611–631. doi:10.3758/BF03197104
- McElreath, R. (2015). *Statistical Rethinking: A Bayesian course with Examples in R and Stan*. Boca Raton, FL: Chapman and Hall/CRC.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 817–835.
- Momma, S., Sakai, H., & Phillips, C. (submitted). *Give me several hundred more milliseconds: Temporal dynamics of verb prediction in Japanese*.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 92–103.
- Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics Series: Structures and processes (Vol. 2)* (pp. 157–187). London: Elek.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from <http://www.usf.edu/FreeAssociation/>
- Nicol, J. (1988). *Coreference processing during sentence comprehension* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. doi:10.7554/eLife.33468
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, 1118(1), 155–167. doi:10.1016/j.brainres.2006.08.022
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 411–421.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Regev, G., & Baddeley, A. D. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, 139, 5–21.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912–927.
- Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22, 642–647.
- Sanford, A. J., & Garrod, S. C. (1989). What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, 4(3–4), S1235–S1262.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382–386.
- Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, 25, 385–401.
- Smith, C., & Federmeier, F. (2015). *N400 modulation profile across sentential contexts informs time course of semantic activation during referential processing*. Poster presented at the 7th Annual Meeting of the Society for the Neurobiology of language, Chicago, IL.
- Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, 12(3), 175–200.
- Stan Development Team. (2017a). RStan: The R interface to Stan. R package version 2.16.2. Retrieved from <http://mc-stan.org>
- Stan Development Team. (2017b). Stan modeling language users guide and reference manual. Version 2.17.0. Retrieved from <http://mc-stan.org/>
- Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology*, 60(12), 1680–1696.
- Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29, 581–595.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. doi:10.1037/0278-7393.31.3.443
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18, 380–393.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. doi:10.1162/0898929041920487