

Towards a Computational Model of Actor-Based Language Comprehension

Phillip M. Alday · Matthias Schlesewsky ·
Ina Bornkessel-Schlesewsky

Published online: 4 August 2013
© Springer Science+Business Media New York 2013

Abstract Neurophysiological data from a range of typologically diverse languages provide evidence for a cross-linguistically valid, actor-based strategy of understanding sentence-level meaning. This strategy seeks to identify the participant primarily responsible for the state of affairs (the actor) as quickly and unambiguously as possible, thus resulting in competition for the actor role when there are multiple candidates. Due to its applicability across languages with vastly different characteristics, we have proposed that the actor strategy may derive from more basic cognitive or neurobiological organizational principles, though it is also shaped by distributional properties of the linguistic input (e.g. the morphosyntactic coding strategies for actors in a given language). Here, we describe an initial computational model of the actor strategy and how it interacts with language-specific properties. Specifically, we contrast two distance metrics derived from the output of the computational model (one weighted and one unweighted) as potential measures of the degree of competition for actorhood by testing how well they predict

modulations of electrophysiological activity engendered by language processing. To this end, we present an EEG study on word order processing in German and use linear mixed-effects models to assess the effect of the various distance metrics. Our results show that a weighted metric, which takes into account the weighting of an actor-identifying feature in the language under consideration outperforms an unweighted distance measure. We conclude that actor competition effects cannot be reduced to feature overlap between multiple sentence participants and thereby to the notion of similarity-based interference, which is prominent in current memory-based models of language processing. Finally, we argue that, in addition to illuminating the underlying neurocognitive mechanisms of actor competition, the present model can form the basis for a more comprehensive, neurobiologically plausible computational model of constructing sentence-level meaning.

Keywords Computational model · Language processing · Emergence · Ambiguity resolution · Actor identification

Parts of the research reported here were supported by the German Research Foundation (grant BO 2471/3-2) and the EEG study was performed while IBS was at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

P. M. Alday (✉) · I. Bornkessel-Schlesewsky
Department of Germanic Linguistics,
University of Marburg,
Deutschhausstr. 3, 35032 Marburg, Germany
e-mail: phillip.alday@staff.uni-marburg.de

M. Schlesewsky
Department of English and Linguistics,
University of Mainz,
Jakob-Welder-Weg 18, 55099 Mainz, Germany
e-mail: schlesew@uni-mainz.de

Introduction

The roughly 6000 languages of the world present a diverse set of grammars and input forms for the single processing mechanism of the human brain. Fundamental differences in word order, different means of encoding different parts of morphosyntax and broad variation in dropped / elided elements are just some of the variation with which the brain must cope; the complexity of a language is matched and exceeded by the complexity of language diversity. In light of the extreme variance between the languages of the world and their respective grammars, less syntax-bound language processing strategies have been proposed.

Neurophysiological data from a range of typologically diverse languages provides evidence for a comprehension and processing heuristic based on the notion of “actor”, the participant primarily responsible for the state of affairs (Bornkessel-Schlesewsky and Schlesewsky 2009). The role of actor, while correlating strongly with certain parts of morphosyntax in some languages, is a language independent construct and is orthogonal to traditional notions of grammar (Bornkessel-Schlesewsky and Schlesewsky 2013a). Here, we present a computational implementation of the heuristic as well as a quantitative comparison with EEG data from an experiment primarily manipulating word order and its related ambiguities. We show that the actor heuristic is not just an interesting, qualitative theoretical construct, but rather a quantifiable and testable model. Indeed, we show that the quantification of the actor heuristic is a reliable, effective predictor of ERP data.

Neurophysiological Model and Language Processing Strategy

Before turning to the computational model that is the focus of the present paper, we will briefly describe the empirical neurocognitive model on which it is based. This discussion will serve primarily to introduce the critical notion of competition for the actor role, which will be central to the computational model to be introduced later. Having introduced actor competition, we will briefly summarize the empirical evidence in support of it.

The Extended Argument Dependency Model (eADM) and Actor-Centered Comprehension

The extended Argument Dependency Model ((e)ADM; Bornkessel 2002; Schlesewsky and Bornkessel 2004; Bornkessel and Schlesewsky 2006; Bornkessel-Schlesewsky and Schlesewsky 2008, 2009, 2013b) is a neurobiologically motivated, neurocognitive model of language comprehension with an explicit focus on cross-linguistic diversity. In other words, the model aims to account for language processing in typologically diverse languages and to explain which aspects of the processing architecture are universal and which are language-specific.

The eADM posits that language processing is organized in a cascaded, hierarchical fashion and proceeds along two major functional-neuroanatomical streams in the brain. One of these, the postero-dorsal stream, engages in time-dependent computations, while the other, the antero-ventral stream, engages in time-independent computations (Bornkessel-Schlesewsky and Schlesewsky 2013b). Time-dependent computation refers to the notion that, in the combination of two elements, A and B, the order in which

they are encountered is crucial for the way in which they are combined. For example, in German, the order in which two noun phrases are encountered in an NP-NP-V sequence changes the likelihood for one being interpreted as the actor argument as opposed to the other. In time-independent computation, by contrast, the order of encountering two elements A and B will not influence the way in which they are combined. For example, a plausibility-based heuristic which computes the most plausible combination of arguments and the verb (e.g. given “apple” and “eat”, the reading that the apple is the undergoer of the eating event rather than the actor) is independent of which element is encountered first. These time-independent computations are implemented in terms of schema unification (see below for a brief description of schemata and the ventral stream). Thus, time-dependent versus time-independent computations could also be described as “sequence-dependent” versus “sequence-independent” operations.

Processing in both streams is organized in a hierarchical manner in accordance with the neurobiological principle of hierarchical processing (Felleman and Van Essen 1991; Rauschecker 1998; Rauschecker and Scott 2009; DeWitt and Rauschecker 2012) and classic assumptions regarding the structure of complex cognitive models (Simon 1962; Newell 1990). This means that, as information flows along the streams, the representations that are processed are assumed to become increasingly complex.¹ In the following, we will refer to the successive points of information processing within the hierarchy as “processing steps” for convenience, though this is clearly a cognitive term that does not directly reflect the underlying neurobiological organization.

In a first step (ignoring preceding aspects of phonological processing and segmentation), the processing system identifies word categories and uses these to build a constituent structure (“syntactic structuring” within the postero-dorsal stream). Crucially, and in contrast to the assumptions of other comprehension models (Friederici 2002; Hagoort 2005; Vosse and Kempen 2000), this structure does not determine sentence interpretation: this is accomplished via a separate mechanism, as we shall see shortly. A second function of category processing in this step is to classify the current input element in terms of its function, e.g. whether it is referential (“nouny”) or predicating (“verby”).

¹Note that, though the model is hierarchically organized, it is not modular in the traditional Fodorian sense (Fodor 1983). Firstly, due to the cascaded nature of processing, a particular processing step need not be fully complete before the next step is initiated. Secondly, from a neurobiological perspective, connections within each pathway are inherently bidirectional such that top-down modulations of information processing are always possible. Nevertheless, we assume that there is an asymmetry in the directionality of information flow based on the tenet of hierarchical organization.

In a second step, sentence-level interpretive mechanisms set in. In the postero-dorsal stream, the system determines sentence meaning from an action-based perspective by assessing who or what is primarily responsible for the state of affairs being described, i.e. here, the actor heuristic mentioned above comes into play. The antero-ventral stream, by contrast, constructs a schema-based representation of sentence-level meaning via the unification of “actor-event schemata”. For reasons of brevity, we will not go into details with respect to the properties of these schemata or their unification, and focus instead on the “actor computation” step posited as part of the postero-dorsal stream. For a detailed discussion of actor-event schemata, the interested reader is referred to Bornkessel-Schlesewsky and Schlewsky (2013b).

The notion of “actor computation” within the postero-dorsal stream is based on the assumption that a linguistic actor is a “stable, language-independent category, possibly rooted in the human ability to understand goal-directed action” (p. 250) (Bornkessel-Schlesewsky and Schlewsky 2013a). The fact that humans are generally attuned to this category as opposed to others could be due to basic evolutionary demands. In the words of Leslie (1995): “Agents are a class of objects possessing sets of causal properties that distinguish them from other physical objects” and “as a result of evolution, we have become adapted to track these sets of properties and to efficiently learn to interpret the behaviour of these objects in specific ways” (p. 122). By tracking (potential) actors, i.e. those entities that appear suited to bringing about changes in the environment (e.g. warranting a fight-or-flight response), we can interpret the world around us and make predictions about upcoming events (see also Frith and Frith 2010). In accordance with this assumption, it has been demonstrated that the human attention system appears to have developed a special sensitivity towards humans and non-human animals (i.e. good potential actors) as opposed to other categories (New et al. 2007). In this way, the actor-centered comprehension strategy posited by the eADM essentially views a sentence as an instruction to conceptualize a particular scenario in which an actor is engaged in a certain event or state of affairs.

How are actor participants identified during language comprehension? In this regard, we have proposed that the prototypical actor may be modeled on the first person (i.e. the self as an acting agent, see Tomasello 2003, Haggard 2008). According to Dahl, this “self-as-actor” perspective is tied to humans perceiving conspecifics as being “like myself, individuals who can perceive the world and act upon it” (Dahl 2008, p. 149). Thus, in order to understand the environment around us, we use the self as a model for other animate entities (and particularly other humans), which in turn serve as a model for inanimate entities. In view of these considerations, the language comprehension

system uses the features +self and +animate as cues to the identification of actor participants (see below for a summary list of actor features). Furthermore, in line with the notion that the self-as-actor perspective involves seeing others as individuals (i.e. other “selves”), an optimal actor is individuated (i.e. definite and specific). Finally, actorhood correlates with particular morphosyntactic features, which are partly cross-linguistically applicable (in particular: occurring as the first argument in a sentence) and partly language-specific (e.g. nominative case marking). Thus, the different features vary in applicability across languages; they also vary in their language-specific weighting, i.e. their importance to identifying actor participants in a particular language (cf. also MacWhinney and Bates 1989; Bates et al. 2001). We have posited that language-particular cues to actorhood (e.g. the importance of morphological case marking in a language such as German) are acquired via their high degree of co-occurrence with the universal actor features based on the first person model (Bornkessel-Schlesewsky and Schlewsky 2013a). Thus, prototypical actor features derived from the self-as-actor perspective are used to bootstrap other, language-specific (morphosyntactic) features of actor participants—a view that is similar to that adopted by emergentist models of other linguistic categories such as parts of speech (Croft 2001).

Linguistic prominence features related to actor identification.

1. +self
2. +animate/+human
3. +definite/+specific
4. +1st position (correlates with actorhood cross-linguistically; (Tomlin 1986))
5. +nominative (correlates with actorhood in nominative-accusative languages with morphological case)

The degree to which arguments are good competitors for the actor role is defined by two points: (a) their own prototypicality in terms of the defining actor features and the correlating prominence features (see above), and (b) the existence and prototypicality of further competitors. Thus, an initial argument is preferentially analyzed as an actor even if it is not highly prototypical (e.g. if it is inanimate).²

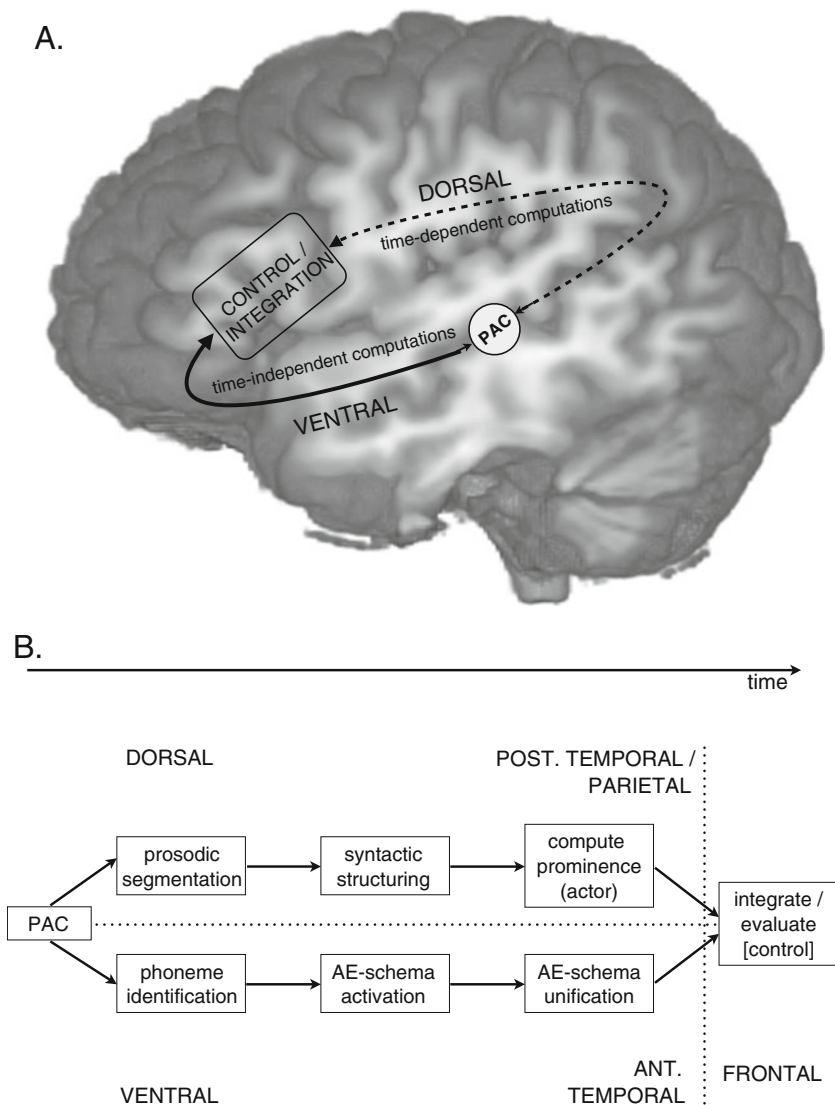
²In this regard, the assumptions of the eADM differ from those of the Competition Model (e.g. Bates et al. 1982, 2001, MacWhinney and Bates 1989), which assumes that a strong cue for the undergoer role (e.g. accusative case in a language such as German or Hungarian) can, to all intents and purposes, exclude an argument from being considered a potential actor. The eADM, by contrast, posits that a sole argument is always considered for the actor role no matter how bad a candidate it is—unless there is a second, more optimal candidate (for discussion, see Bornkessel-Schlesewsky and Schlewsky to appear).

Once a second argument—a competitor—is encountered, however, the relative actor prototypicality of the arguments is crucial in determining whether the actor preference for the first argument can be maintained or whether it needs to be revised.

In accordance with the model architecture in Fig. 1, increased competition for the actor role (including the need to revise a previous actor choice) correlates with increased activation in the posterior superior temporal sulcus and the temporo-parietal junction as part of the postero-dorsal stream (Bornkessel et al. 2005; Grewe et al. 2006; Bornkessel-Schlesewsky and Schlesewsky 2009) and the anterior temporal lobe as part of the antero-ventral stream (Magnusdottir et al. 2012; for discussion, see Bornkessel-Schlesewsky and Schlesewsky 2013b). In neurophysiological terms, it is reflected in increased amplitude of the N400 event-related brain potential (Bornkessel-Schlesewsky and Schlesewsky 2009, 2013a). We assume that the N400—

and negative ERP deflections in general—result from a mismatch between top-down and bottom-up information sources within the two processing streams (for proposals that the N400 depends on an integration of top-down expectations and bottom-up input, see (Federmeier 2007; Lotze et al. 2011)). Crucially, as scalp ERPs are macroscopic responses which typically result from the mixing of multiple underlying sources, the claim is not that an N400 effect elicited by actor competition results from activation changes in only a single locus within a stream. Rather, it is likely due to the summed reaction of top-down/bottom-up integration within multiple processing steps within both streams (e.g. “actor computation” within the postero-dorsal stream and “actor-event schema unification” within the antero-ventral stream). In addition, since we posit that the notion of top-down/bottom-up integration can be generalized to other language-related negativities (e.g. left-anterior negativity (LAN) effects, which result, for example, from

Fig. 1 Model architecture for the latest version of the extended Argument Dependency Model, eADM (adapted from (Bornkessel-Schlesewsky and Schlesewsky 2013b)). Panel A provides a basic overview of the model’s neuroanatomical assumptions: the ventral (*solid line*) and dorsal (*dashed line*) streams are assumed to emanate from primary auditory cortex (PAC) and to perform information processing in a hierarchically organized manner. Thus, in spite of the fact that the streams are inherently bidirectional, there is an asymmetry in the directionality of information flow on account of the hierarchical organization. Panel B shows the assumed structure of hierarchical processing within the two streams



subject-verb agreement errors in which the expectation for a particular agreement morpheme is not met), latency and topography of the negativity response are assumed to vary depending on the loci within the streams giving rise to the mismatch and the timing of their activation (Sassenhagen et al. 2013).

When actor competition is behaviorally relevant (e.g. when participants perform an acceptability judgment task and actor competition affects how acceptable a sentence is deemed to be), it additionally engenders a late positive ERP response. In this view, late positivities in language processing (“P600” effects) are viewed as members of the domain-general P300 family (Coulson et al. 1998; Roehm et al. 2007; Kretzschmar 2010). Recently, Sassenhagen et al. (2013) linked this idea to a neurobiological model of the P300, the Locus Coeruleus-Norepinephrine (LC/NE-P3) model (Nieuwenhuis et al. 2005). According to this model, the P300 results from activation of the Locus Coeruleus (i.e. the brain stem source for noradrenergic projections to the cortex) following the detection of subjectively significant events. This results in a release of norepinephrine, thereby increasing neural responsivity to a particular stimulus and influencing the behavioral response to it. From this perspective, late positivities in language processing reflect a systemic neuromodulator release facilitating the application of decision processes rather than linguistic processing per se. In support of this view, Sassenhagen et al. (2013) found that an N400 – late positivity scalp ERP pattern engendered by semantically anomalous versus plausible words in a highly constrained sentence context could be decomposed using independent component analysis (ICA) and single-trial analyses. Results revealed that, while N400 effects were timelocked to critical stimulus onset, positivity effects were response-locked.

In summary, actor competition is reflected in N400–late positivity patterns in electrophysiological studies, though, as described above, the two components of this biphasic response are functionally distinct and, in principle, independent of one another. In addition, the presence or absence of the late positivity effect is conditioned by the experimental environment and task.

Evidence for the Actor Heuristic and for Competition for the Actor Role

The eADM’s notion of competition for the actor role is supported by a range of cross-linguistic studies on sentence comprehension, which have provided evidence for the following generalization regarding online-processing:

Cross-linguistic generalization regarding actor identification in online language processing (Bornkessel-Schlesewsky and Schlesewsky 2009):

The processing system attempts to identify the actor role as quickly and unambiguously as possible.

Corollaries:

- The processing system prefers actor-initial orders
- The processing system prefers prototypical actors.

Evidence for this generalization stems from electrophysiological studies in a number of typologically varied languages, including Turkish (Demiral et al. 2008), Chinese (Wang et al. 2009) and Hindi (Choudhary et al. 2010), thus corroborating previous findings of a “subject-first preference” in European languages (e.g. Dutch: Frazier (1987), German: Schriefers et al. (1995), Schlesewsky et al. (2000), Bornkessel et al. (2004b), amongst others; Italian: de Vincenzi (1991), Penolazzi et al. (2005)). Importantly, the empirical findings from non-Indo-European languages (Chinese and Turkish) support the assumption of an actor-first rather than a subject-first preference, since they rule out explanations based solely on formal subject features such as agreement. They further suggest that the actor-first preference cannot be reduced to structural simplicity or frequency (see Wang et al. (2009) for a summary). The finding of an actor-first preference even in an ergative language (Hindi) further demonstrates the need to assume an actor-first as opposed to a subject-first preference.³

The preference for prototypical actors shows up in a similarly ubiquitous way: When an argument that is unambiguously the actor in a transitive (two participant) relation is non-prototypical because it is inanimate, different languages consistently show an N400 effect (for a comprehensive review, see (Bornkessel-Schlesewsky and Schlesewsky 2009)). For an illustration, consider the following example from Frisch and Schlesewsky (2001):

- (1) a. Paul fragt sich, welchen Förster
Paul asks himself, [which forester]_{ACC}
der Zweig ...
[the twig]_{NOM} ...
- b. Paul fragt sich, welchen Angler
Paul asks himself, [which angler]_{ACC}
der Jäger ...
[the hunter]_{NOM} ...

³In an ergative language such as Hindi, the actor argument in a transitive (two-participant) event is not morphosyntactically “privileged” in the sense that it does not agree with the verb, for example. Thus, it does not qualify for grammatical subjecthood in the same way as a transitive actor in a non-ergative language such as German, Dutch or Italian. The results from Hindi thus provide strong converging support for the assumption that the actor preference is interpretive rather than grammatical in nature.

In example (1), the initial accusative—as a very poor actor candidate—leads the processing system to expect to encounter a better actor candidate as a second argument. When this expectation is contradicted by the features of an inanimate second argument (1a), which is also an atypical actor, an N400 effect arises in comparison to a control condition with an animate second argument (1b). Thus, as described in the preceding section, the N400 effects arises from a mismatch between top-down information (the expectation) and bottom-up information (the features of the second argument). In addition to German (Frisch and Schlesewsky 2001; Roehm et al. 2004) and English (Weckerly and Kutas 1999), this effect has been shown in Mandarin Chinese (Philipp et al. 2008) and Tamil (Muralikrishnan et al. 2008).

Why a Computational Model of Actor Competition?

As is apparent from the preceding section, the actor-based comprehension strategy is well described in qualitative terms. In formalizing this strategy with an implemented model, we pursue a twofold aim. Firstly, from the computational implementation of the actor strategy, we aim to gain quantitative predictions that can be tested against empirical data. This will allow for the predictions of the eADM to be tested in a more stringent manner and for cross-linguistic similarities and differences to be expressed in more explicit terms.

Secondly, these quantitative predictions can be used to illuminate the basic processing mechanisms underlying the actor strategy. In particular, we aimed to compare two alternative conceptualizations of actor competition: unweighted similarity-based interference and weighted competition. Similarity-based interference is a notion that features prominently in contemporary approaches to working memory (WM), which emphasize the status of WM as the activated portion of long-term memory rather than as a separate buffer (for overviews, see McElree (2006), Jonides et al. (2008); for approaches to language-processing based on this notion, see Lewis et al. (2006), Lewis and Vasishth (2005)). Accordingly, memory retrieval is conceptualized not as the result of a (serial) search, but of a content-addressable pointer mechanism based on so-called retrieval cues. These cues (for example, case, number or other features) provide the relevant information required to access the item in question. Retrieval becomes more effortful when cues overlap (i.e. apply to several items in memory), a phenomenon termed “similarity-based interference”.

Similarity-based interference appears well suited as a potential mechanism underlying actor competition effects

(Bornkessel-Schlesewsky and Schlesewsky 2013a), which, as described in detail above, arise when multiple candidates within a clause bear actor features. Since interference in its typical form (i.e. as conceptualized within the WM literature) is based exclusively on feature overlap, it predicts that the degree of competition for the actor role should be a function of the number of actor features shared by the arguments, while the specific weighting of a feature within a language should be irrelevant. More directly: feature overlap is an all-or-nothing measure for individual features because either two entities overlap in a given feature or not. By contrast, a second potential conceptualization of the degree of actor competition is that it goes beyond similarity-based interference in the classical sense and rather also takes into account the importance of a particular feature for actor recognition in the language under consideration. From this perspective, the degree of actor competition should be proportional to the difference in prominence (i.e. goodness of fit to the actor role) between arguments with individual features weighted according to their language-specific importance.

By means of the computational implementation introduced in the next section, we will calculate explicit metrics for the two alternative conceptions of actor competition outlined above and will test these against data from a neurophysiological experiment on sentence processing.

In this way, we endeavor to use the computational implementation of the actor strategy not only as a means of deriving more precise (quantifiable) predictions, but also to shed further light on the how the strategy is neurocognitively implemented.

Computational Implementation

The present implementation focuses on the core calculation of actor competition for referential elements, called Compute Prominence in previous versions of the eADM (Bornkessel-Schlesewsky and Schlesewsky 2009; Bornkessel and Schlesewsky 2006). For convenience, we similarly use the existing terminology “Stage 1” to refer to the initial chunking and analysis step and “Stage 2” to refer to the sentence-level interpretative mechanisms of the second step. A brief summary of the current software implementation can be found in the Technical Notes at the end of this article (p. 29).

The computational implementation does impose one restriction that the neurocognitive model upon which it is based does not: Stage 1 completes in full before Stage 2 begins. This is however not as detrimental to the approximation as it may initially seem because Stage 2 processes each constituent incrementally in the original sequential

order. Furthermore, both incremental and final full processing results for Compute Prominence are computed and optionally displayed.⁴

The completion of Stage 1 in its entirety before Stage 2 is unfortunately not capable of modeling cases where additional, disambiguating information becomes available. In German, this primarily happens in noun phrases via gender (indefinite NPs) and number (definite NPs) information available on the first non-article adjective or head noun. (Disambiguation via verb agreement is also possible, but this is an interaction with the computation for predicating elements—Compute Linking in previous versions of the eADM—and is not currently modeled for non verb-final word orders.) However, none of these forms of disambiguation occur in the present experimental manipulation. Nonetheless, processing of ambiguities remains a major focus of present and future research.

Stage 1

The initial chunking and morphological analysis in Stage 1 is performed here only in a restricted fashion. The full complexity of German phrase structure would be a non-trivial undertaking and lies outside the scope of this paper and its focus on the actor heuristic (Stage 2) and an appropriate computational implementation. However, a sufficient implementation of Stage 1 to parse the stimuli from an EEG/ERP experiment with their relatively rigid structure is possible.

In the present experiment (see section “[EEG Experiment](#)”), it suffices to process inflection carried via pronouns and articles. In German, the article carries the majority of the morphological burden in noun phrases. The head noun inflects for number and can carry an additional marker for dative in the plural; however, this information was redundant in the present experiment, where neither dative nor plurals were used.⁵ In the pronominal system, there exist a few ambiguities, especially between the nominative and accusative 3rd person for neutra and feminina as well as in the plural. This ambiguity could potentially be resolved by agreement with the verb; however, it also presents a general test case for the heuristic implemented by Compute Prominence. The ambiguity is thus marked as such and otherwise

not further processed by Stage 1. The further ambiguity in the pronominal system between the 3rd person feminine dative and the second person plural nominative is always resolved by verbal agreement, but as there are no datives in this experiment, this special case is not processed further and is implemented by `pass` (a syntactic placeholder in Python similar to `void`) in the branch construct.

Stage 2

Implementation of Stage 2 was restricted to the function Compute Prominence, which provides in non-headmarking languages the most important parts of the actor heuristic in single-sentence processing.⁶ Furthermore, Compute Prominence remains largely unchanged in recent and further planned updates to the eADM.

In implementing Compute Prominence, we view the hierarchies as dimensions in (a subspace of) Euclidean n -space, where n is the number of hierarchies. The prominence of an individual argument is thus a vector, with each component being a scalar representing the prominence with respect to a particular hierarchy. The hierarchies with the corresponding values for various linguistic features in the current implementation are given in Table 1. The “additional” feature NUMBER derives from another prototypical feature of the self-as-actor view: singular correlates with stronger individuation. Negative values are used to actively penalize a particular prominence component in the next calculation. That is, negative values indicate a feature that strongly correlates with a poor actor candidate (designated in the computational model by the feature `+dep`).⁷

For now, we make the a priori assumption that case is a singular feature with multiple levels cf. (Kempe and MacWhinney 1999). However, it is possible that “case” is merely a convenient moniker for a set of strongly correlating binary features such as \pm nom and \pm acc. Ambiguity would then be encoded by setting all individual case features to the same value, e.g. `+acc`, `+nom` for an ambiguity between nominative and accusative. This latter approach is the typical one found in NLP and has the interesting feature that individual cases can carry different, individual weights (see below). For example, `+acc` may be a much stronger indicator of a particular role assignment than `+nom`.

⁴This restriction exists primarily to simplify the implementation in a single Python program (see Technical Notes); to better model the waterfall data flow, coroutines could be used or Stage 1 and Stage 2 could be split into two programs connected by Unix pipes.

⁵Furthermore, the article also carries number information, albeit with a small ambiguity that is resolved through further adjectives or the marking on the head noun.

⁶Of course, contextual effects also play a role in normal language use.

⁷For the purposes of the present paper, `+dep` may simply be considered a convenient label for a poor actor candidate. For an in-depth discussion of \pm dep and a motivation in terms of a previous version of the eADM, see (Bornkessel and Schlesewsky 2006)

Table 1 Prominence hierarchies and the corresponding scalar values for the various features as used in the current implementation of Compute Prominence

Feature	Hierarchy
Person:	First = 1 > Other = 0
Case:	Nominative = 1 > Dative = 0 > Accusative = -1
Animacy:	Animate = 1 > Inanimate = 0
Position:	Early = 1 > Late = 0
Number:	Singular = 1 > Plural = 0
Definiteness:	Definite = 1 > Indefinite = 0

Similarly, other features with multiple levels are encoded binarily according to their most prominent tendency.

\pm PERSON is actually \pm IST. PERSON and not a multi-tiered variable. Extending this to accommodate the second person would only require the addition of a further field in the prominence vector. As all functions in the implementation are written to handle vectors of arbitrary length, this would require no changes to the core code. However, the question remains open for the model development, whether representation as multiple fields or as a variable with more degrees of freedom is the sensible choice. Multiple variables allow for learning the weights (and hence the impact) of distinct levels separately; however, this potentially allows for unlikely combinations of multiple levels.⁸ A final consideration in the weight encoding is the use of fuzzy logic for the boolean values. For example, an ambiguous noun phrase could be assigned values between zero and one to indicate some form of probability for a given analysis. A plant could be assigned an animacy value of 0.25 (alive and able to die but largely not capable of independent action), an animal could be assigned a value of 0.8 (alive and capable of independent action, but not sentient) and a human a value of 1.0 (alive, willful and sentient). Furthermore, this corresponds with animacy hierarchies seen in the languages of the world, with similar ordering, but language-specific cut-off points between levels (Silverstein 1976). Such gradience adds a flexibility to the use of binary features at the cost of making the prominence encoding somewhat less sparse.

The language-specific relative weights are also stored in a vector in the same space. The scalar (dot) product of the weight vector with the prominence vector yields a scalar value for the total prominence. This value is then compared with a threshold value to determine if +dep is assigned immediately. Compute Prominence is applied to

⁸This is perhaps an advantage—in languages where inclusive and exclusive first person are morphologically distinct, this could be represented by the interaction of \pm 1st. Person and \pm 2nd. Person. This added complexity nonetheless introduces its own cost and brings language specific features deeper into the model.

both arguments and the values are compared, with the more prominent argument being assigned $-dep$ (i.e. designated as the actor argument) and the less prominent argument +dep.⁹ In the special case of an Object-Experiencer verb, prominence values for the case hierarchy are inverted: the entire hierarchy is multiplied by -1, thus reversing the orientation such that the accusative and dative outrank the nominative.

A sample sentence set for the EEG experiment used here as well as an analysis for a single condition of that set are given in Tables 2 and 3.

Distinctness/Actor Competition

The use of vectors to represent prominence data also allows for several other calculations to be made. The magnitude of the projection of an argument's prominence vector on the prominence vector for an idealized actor or undergoer is an index for the prototypicality of a particular argument. Similarly, the scalar product of the two argument vectors corresponds inversely with distinctness.

Distinctness is more broadly the distance between two arguments in actor-space. It thus provides a measure for the degree of competition for the actor role: when distinctness is low, multiple arguments bear actor features and competition for the actor role is high; when distinctness is high, actor features accumulate on only a single argument and competition for the actor role is low. Various metrics are provided (selectable as command line options in the implementation here) for the distance measurement. The Manhattan metric¹⁰ reflects the summed distance between individual features. This is the default in the model and reflects an intuitive notion of distinctness. Furthermore, the Manhattan metric provides a general measurement of feature overlap and thus correlates inversely with traditional notions of interference—the fewer features that overlap/interfere, the larger the Manhattan distance. The Euclidean metric¹¹ is also provided and reflects a more continuous notion of distinctness. Finally, the difference in Euclidean magnitude of the two vectors is provided as a metric reflecting the difference in “absolute” (unweighted scalar) prominence. This magnitude difference also reflects the directionality of the prominence shift—the prominence of NP1 is subtracted from the prominence of NP2. Thus, NP2 is more prominent if and only if the magnitude difference is

⁹In the case of a single argument, e.g. intransitivity, the distinction measure is not performed and the model depends solely on the threshold comparison. The present experiment included only monotransitive sentences.

¹⁰ $d(x, y) = \sum_i |y_i - x_i|$

¹¹ $d(x, y) = \sqrt{\sum_i (y_i - x_i)^2}$

positive. The selected metric is outputted (in batch mode; see Technical Notes) as the field `dist`. For the data here, this is the Manhattan metric, and so `dist` is equal to $\sum_i |NP2_i - NP1_i|$.

Distinctness as distance in actor space is calculated on the raw prominence vectors without the weight distortion. This has the simultaneous advantage and disadvantage that the distance is language independent.¹² Weighted distance is given by the field `sdiff` (in the batch mode output) and is calculated as the difference of the weighted, scalar prominences: $\vec{w} \cdot N\vec{P}2 - \vec{w} \cdot N\vec{P}1 = \sum_i w_i \cdot NP2_i - \sum_i w_i \cdot NP1_i$, where \vec{w} is the language specific weight vector. Since the weighted scalar prominences are calculated by the dot product of the weight and feature vectors, this is equivalent to the weighted, signed Manhattan distance.¹³ Baseline weights were based on previous work done in German in the framework of the Competition Model cf. (Kempe and MacWhinney 1999; MacWhinney et al. 1984). The qualitative ordinal scales were converted to quantitative interval scales via a simple order of magnitude mapping: for a feature f_1 ranked more strongly than another feature f_2 , $f_1 = 10f_2$ (see Stevens (1951) for the classification of scales). If a feature is considered a much stronger cue than another, then two orders of magnitude of separation was assumed: $f_1 = 100f_2$. Thus, we have the following ordering: *case* = 1000 > *position*, *person* = 100 > *animacy*, *number* = 10 > *definiteness* = 1.¹⁴ The exact numerical values have no empirical meaning in their own right. Rather

¹²In as far as all features are treated equally—some languages may not take advantage of certain features, e.g. English largely does not use case.

¹³This follows very straightforwardly from the definition and standard properties of the dot product:

$$\begin{aligned} \vec{w} \cdot N\vec{P}2 - \vec{w} \cdot N\vec{P}1 &= \sum_i w_i \cdot NP2_i - \sum_i w_i \cdot NP1_i \\ &= \sum_i w_i \cdot (NP2_i - NP1_i) \\ &= \vec{w} \cdot (N\vec{P}2 - N\vec{P}1) \end{aligned}$$

¹⁴As noted by an anonymous reviewer, the EEG experiment presented below does not include any number or animacy contrasts. This however does not present any great problem for the data at hand: due to properties of the metrics at hand, the non contrasting features simply cancel out and do not even introduce additional parametric levels into the respective prominence metrics. These features remain in the models present because their presence does not detract from the comparisons in question and avoids an experiment-specific model. One subtle disadvantage does come into play here though: the fit of the weights for these two features is not tested. Especially our ranking of position relative to animacy may prove problematic and, as such, more explicit testing, manipulation and determination of model weights is planned for future research.

their relationship to one another is central.¹⁵ Changing the precise values will of course change the coefficients in the fitted mixed linear model, but will not change the properties of the model as a whole.¹⁶

In terms of the research questions introduced above, the main aim of the present paper is to compare the distinctness measures `dist` and `sdiff` as predictors of empirical neurophysiological data. While `dist` provides a good measure of similarity-based interference, `sdiff` implements the alternative, weighted notion of distinctness.

EEG Experiment

We tested the effectiveness of the model parameters as predictors of neurophysiological activity using data from an EEG experiment on word order processing in German.

The experiment manipulated actor competition by varying actor-undergoer order and case-marking ambiguity in transitive sentences with a noun phrase (NP1) – verb – noun phrase (NP2) structure. In particular, it examined sentences which—due to locally ambiguous case information—were initially compatible with an actor-first reading but subsequently required a reinterpretation as undergoer-initial. These are cases where actor competition is particularly high. They were compared with locally ambiguous sentences in which the actor-first preference was borne out and competition for the actor role was thus considerably less pronounced as well as with unambiguously case-marked sentences. Crucially for present purposes, the relative prominence of the two arguments—and hence their relative degree of actor prototypicality—was also manipulated in order to induce more subtle variations of actor competition. To this end, NP1 was either realized as a non-pronominal NP or as a 3rd person pronoun and NP2 was either realized as a non-pronominal NP or a 1st person pronoun. Recall from section “The Extended Argument Dependency Model (eADM) and Actor-Centered Comprehension” that optimal actorhood is assumed to be modeled on the first person within

¹⁵This follows from the notion of an actor space—we can expand or contract the space by a constant multiple without changing the inherent properties of it. Specifically, $c\vec{v} \cdot c\vec{w} = c(\vec{v} \cdot \vec{w})$.

¹⁶Subject to the constraints of the effects this has on precision and representation on the computing machine in question. Theoretically, we could divide all of these values by 1000 (the maximum weight given here), giving us coefficients on [0, 1], which would reflect their impact in the notation of probability theory. This is a very interesting approach, as the deterministic impact of case would receive a (probability) coefficient of one—certainty. However, this all too easily leads to the assumption that there is necessarily a single feature which, when unambiguous, is singularly deterministic in its influence. Or, in the particular case of German, that the impact of case is always deterministic—clearly, this is not the case as all too often, the morphological marking is ambiguous: 0×1000 is still 0.

Table 2 Stimulus design in the EEG experiment. Every condition appeared for each lexical item

Initial	Ambiguous	NP1-Type	NP2-Type	Sentence(Example)
Actor	Yes	Noun	Noun	Die Bettlerin bedrängte den Kommissar auf der Straße.
Undergoer	Yes	Noun	Noun	Die Bettlerin bedrängte der Kommissar auf der Straße.
Actor	No	Noun	Noun	Der Bettler bedrängte den Kommissar auf der Straße.
Undergoer	No	Noun	Noun	Den Bettler bedrängte der Kommissar auf der Straße.
Actor	Yes	Noun	Pronoun (1sg)	Die Bettlerin bedrängte mich auf der Straße.
Undergoer	Yes	Noun	Pronoun (1sg)	Die Bettlerin bedrängte ich auf der Straße.
Actor	No	Noun	Pronoun (1sg)	Der Bettler bedrängte mich auf der Straße.
Undergoer	No	Noun	Pronoun (1sg)	Den Bettler bedrängte ich auf der Straße.
Actor	Yes	Pronoun (3sg)	Noun	Sie bedrängte den Kommissar auf der Straße.
Undergoer	Yes	Pronoun (3sg)	Noun	Sie bedrängte der Kommissar auf der Straße.
Actor	No	Pronoun (3sg)	Noun	Er bedrängte den Kommissar auf der Straße.
Undergoer	No	Pronoun (3sg)	Noun	Ihn bedrängte der Kommissar auf der Straße.
Actor	Yes	Pronoun (3sg)	Pronoun (1sg)	Sie bedrängte mich auf der Straße.
Undergoer	Yes	Pronoun (3sg)	Pronoun (1sg)	Sie bedrängte ich auf der Straße.
Actor	No	Pronoun (3sg)	Pronoun (1sg)	Er bedrängte mich auf der Straße.
Undergoer	No	Pronoun (3sg)	Pronoun (1sg)	Ihn bedrängte ich auf der Straße.

The base sentence (first example) translates to “The beggar hassled the commissioner in the street.” The gender of NP was varied for the ambiguity condition; the person of NP2 was varied for the NP2-type condition. Abbreviations: 3sg = third person singular, 1sg = first person singular

the eADM. Accordingly, 1st person pronouns are optimal actors, 3rd person pronouns (which are not 1st person, but nevertheless highly individuated) are somewhat less optimal actors and non-pronominal noun phrases are somewhat less optimal again. By manipulating person rather than more commonly examined actor features such as animacy, the present study therefore allowed us to test the effectiveness of our computational implementation of actor computation (Compute Prominence) as well as the self-as-actor perspective.

Participants

Thirty-seven monolingually raised native speakers of German (20 women; mean age: 25.9 years, range: 20–40 years) participated in the EEG study after giving written informed consent. Participants were right-handed as assessed by a German version of the Edinburgh handedness inventory (Oldfield 1971). The majority of the participants were students at the Free University Berlin at the time of the experiment. Two additional participants were excluded due to technical problems or a failure to complete both experimental sessions.

Materials

The critical sentence types used in this study are shown in Table 2. Sixty sets of the conditions shown in Table 2 were constructed, thus resulting in a total of 960 critical

sentences. These were subdivided into two lists of 480 sentences each (30 from each condition and 8 from each lexical set). The critical sentences for each list were pseudo-randomly interspersed with 240 filler sentences. Fillers

Table 3 Summarized analysis for the sentence *Die Bettlerin bedrängte den Kommissar auf der Straße* “The beggar hassled the commissioner in the street”

Feature	NP1	NP2	Weight
Case	0	−1	1000
Animacy	1	1	10
Person	0	0	100
Number	1	1	10
Definiteness	1	1	1
Position	1	0	100
Prominence			
Simple	5	2	
Weighted	121.0	−979.0	
Metrics			
dist	2		
signdist	+2		
sdiff	−1100.0		

Please note that the order of operation in computing the metrics matters: sum the pairwise differences (with absolute values, no weighting, or weighting, for dist, signdist and sdiff, respectively)

were also declarative main clauses of German but did not contain case or word order ambiguities. Eighty of the filler sentences were ungrammatical due to a case or agreement violation and 60 were semantically implausible, thus ensuring that participants needed to take into account both the grammaticality of the sentences and their plausibility when performing the acceptability judgement task (see below). The filler sentences were the same across the two lists. List presentation was counterbalanced across participants, with each participant reading the sentences from one list once.

Procedure

Participants were seated in a dimly lit, sound-attenuated booth, approximately 1 meter in front of a 17 inch computer screen. Sentences were presented visually in a phrase-by-phrase manner (i.e. noun phrases were presented together as chunks). Each trial began with the presentation of a fixation asterisk (presentation time: 300 ms, followed by an inter-stimulus-interval, ISI, of 200 ms). Single words were presented for 400 ms and phrases for 500 ms, with an ISI of 100 ms in each case. Following the presentation of the sentence-final word or phrase, there were 500 ms of blank screen, after which a question mark signalled to participants that they should judge the acceptability of the preceding sentence using two hand-held push-buttons. They were instructed that their judgement should be based both on form and content (i.e. also take into account the plausibility of the sentence). Assignments of the left and right buttons to “yes” and “no” responses were counterbalanced across participants. Following the judgement or after the maximal reaction time of 2000 ms had expired, there was an inter-trial interval (blank screen) of 1000 ms before presentation of the next sentence began. The experiment was conducted in two sessions, separated by approximately a week. In each session, a participant read 8 blocks of 45 sentences each, with blocks separated by short breaks. Sessions lasted approximately 3 hours including electrode preparation.

EEG Recording and Preprocessing

The EEG was recorded from 25 Ag/AgCl electrodes fixed at the scalp by means of an elastic cap (Easy Cap GmbH, Herrsching, Germany). AFZ served as ground. Electrodes were positioned according to the international 10-10 system. The electrooculogram was monitored by means of electrodes placed at the outer canthi of both eyes (horizontal EOG) and above and below the participant’s right eye (vertical EOG). EEG and EOG channels were amplified by means of a Refa amplifier (Twente Medical Systems, Enschede, The Netherlands) and digitized with a sampling rate of 250 Hz. Channels were referenced to the left mastoid but

rereferenced to linked mastoids offline. In order to eliminate slow signal drifts, a 0.3–20 Hz band-pass filter was applied to the raw EEG data. Trials containing EEG or EOG artifacts were excluded from the final data analysis (the EOG rejection criterion was 40 μV). For display purposes only, the grand average ERPs were smoothed with an 8Hz low-pass filter.

EEG Data Analysis

In an initial step, we performed a standard data analysis for language-related event-related brain potential (ERP) studies. Thus, average ERPs were calculated per condition, electrode and participant from the onset of the critical second noun phrase to 1000 ms post onset, before grand averages were computed over all participants. We then computed a repeated-measures ANOVA with the factors word order (actor-initial versus undergoer-initial), ambiguity (NP1 ambiguous between actor and undergoer versus unambiguously marked), NP1-Type (definite noun phrase versus 3rd person pronoun), NP2-Type (definite noun phrase versus 1st person pronoun) and region of interest (ROI). Lateral regions of interest were defined as follows: left-anterior (F3, F7, FC1, FC5); left-posterior (CP1,

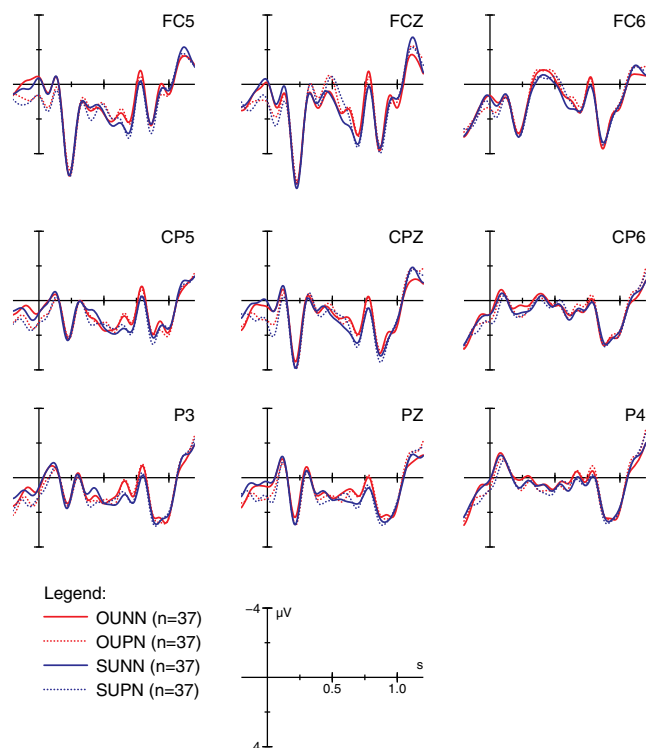


Fig. 2 Grand average ERPs triggered at the onset of NP2 for the unambiguous condition and NP2 a noun with definite article. The condition codes reflect the $2 \times 2 \times 2 \times 2$ design: *S* = subject (actor) initial word order, *O* = object (undergoer) initial; *U* = unambiguous, *A* = ambiguous; *N* = Noun, *P* = pronoun, for NP1 & NP2 respectively

CP5, P3, P7); right-anterior (F4, F8, FC2, FC6); right-posterior (CP2, CP6, P4, P8). A single ROI was used for the midline sites (FZ, FCZ, CZ, CPZ, PZ). For analyses involving more than one degree of freedom in the numerator, significance values were corrected when sphericity was violated (Huynh and Feldt 1970). This analysis was used to identify regions in which the effects were most pronounced for the subsequent analysis using linear mixed effects models, in which we tested the effectivity of the distinctness metrics as predictors of language-related electrophysiological activity.

Linear mixed effects models provide a tool capable of handling the random variation introduced by intersubject differences and lexical effects (Baayen et al. 2008), which are not modeled in the current implementation. Furthermore, they allow continuous predictors such as the actor metrics here, while ANOVA-based analyses do not. Using the R package `lme4` (Bates et al. 2013), we calculated models using subject and item as random factors, and the various distinctness measures as fixed factors. For the random factors, we used the maximal random-effect structure common to all models, i.e. random slopes grouped per distinctness measure, as models without random slopes are anti-conservative (Barr et al. 2013).^{17,18} As an exact estimation of *p*-values in mixed effects models is not straightforwardly possible due to difficulties in estimating the degrees of freedom, we follow Baayen et al. (2008) in considering an absolute *t*-value exceeding 2 as an indication of significance at the 5 %- level.

Intermodel comparisons are also not completely straightforward, especially in the case of non-nested models. Most importantly, log-likelihood tests and the associated χ^2 -statistic (i.e. the parallel to traditional ANOVA, even called via the function `anova()` in R) are only valid for nested models.¹⁹ To compare non-nested models, we turn to information-theoretic criteria (cf. Burnham and Anderson 2002, p. 88). In particular, Akaike Information Criterion (AIC,

¹⁷The models resolved for ambiguity in the P600 time window have only random intercepts, as models with random slopes failed to converge.

¹⁸Higher order interactions were excluded for three reasons. First, comparing models which differ in random-effect structure is less straightforward than those which differ in only fixed-effect structure. (Even for the fixed effects, the comparison between non nested models requires information-theoretic criteria, see main text.) Second, models with higher order interactions in the random-effects structure did not always converge and due to the aforementioned complexities of comparing random-effects structures, it is not clear which of several higher-order models to choose from. Finally, computational complexity increases extremely quickly with random effect complexity. Limiting the random-effects structure to the maximal common one provides an acceptable balance between estimation accuracy, ease of comparison, and computer time.

¹⁹ Models in which the parameters for one model form a proper subset for the parameters of the other.

(Akaike 1974)) and Bayesian Information Criterion (BIC, (Schwarz 1978)) provide further tools for comparing models, based on log-likelihood (fit) penalized by the number of parameters (overfitting). The absolute value of these scores is not meaningful in itself, but the general rule when comparing two models is “smaller is better”. In the following, AIC and BIC are shown in the model summaries, while only AIC with log-likelihood and degrees of freedom for the fixed factor is shown in the model comparisons, since the comparison of non-nested models always involved models with the same number of parameters. For nested models, χ^2 -statistics (based on likelihood ratio tests) are also shown.

Model Performance and Prediction: Results

Behavioral Data

The results of the rating task showed that participants judged all conditions to be highly acceptable (lowest mean acceptability ratings were 86 % for the ambiguous, undergoer-initial condition with two non-pronominal noun

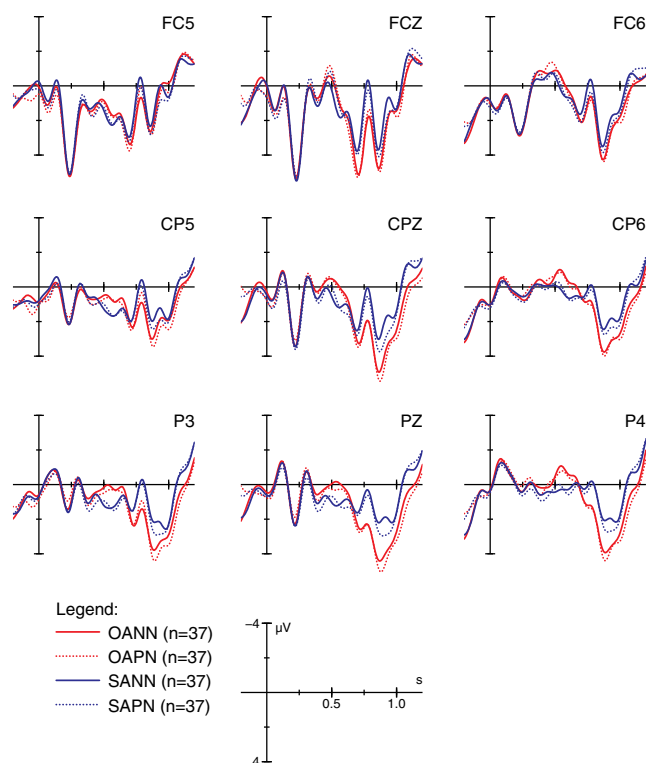


Fig. 3 Grand average ERPs triggered at the onset of NP2 for the ambiguous condition and NP2 a noun with definite article. The condition codes reflect the $2 \times 2 \times 2 \times 2$ design: S = subject (actor) initial word order, O = object (undergoer) initial; U = unambiguous, A = ambiguous; N = Noun, P = pronoun, for NP1 & NP2 respectively

phrases and 89 % for the ambiguous, undergoer-initial condition with NP1 a pronoun and NP2 a non-pronominal noun phrase; all other conditions showed an acceptability of 93 % or higher). We refrain from analyzing the ratings statistically in order to avoid interpreting ceiling effects. Most importantly for present purposes, they demonstrate that participants found the sentences acceptable and that they were able to correctly reanalyze the ambiguous undergoer-initial sentences (which should otherwise have been judged as unacceptable).

The analysis of the reaction times (restricted to sentences correctly judged as “acceptable”) revealed an interaction of AMBIGUITY, WORD-ORDER and NP2-TYPE ($F(1, 36) = 4.95, p < 0.03$) and an interaction of AMBIGUITY and NP1-TYPE ($F(1, 36) = 8.45, p < 0.006$). Resolving both interactions by ambiguity showed an WORD-ORDER \times NP2-TYPE interaction only for locally ambiguous ($F(1, 36) = 11.58, p < 0.002$) but not for unambiguous sentences ($p > 0.15$). For the ambiguous sentences, the interaction was due to longer reaction times for undergoer-initial as opposed to actor-initial sentences when the second noun phrase was non-pronominal / third person (mean RTs of 473 ms for undergoer-initial and 439 ms for actor-initial sentences; $F(1, 36) = 16, 77, p < 0.0003$), while there was no effect of actor-undergoer order when the

second noun phrase was a first person pronoun ($p > 0.22$). The interaction of AMBIGUITY and NP1-TYPE was due to longer RTs for unambiguous sentences in which NP1 was realized as a non-pronominal NP as opposed to a first person pronoun (mean RTs of 443 ms and 431 ms, respectively; $F(1, 36) = 8.53, p < 0.006$).

In summary, reaction times were longer when sentences required a reanalysis towards an undergoer-initial order—as expected from the perspective of an actor-first preference—but this effect was only observable when the disambiguating second noun phrase was a non-pronominal third person, not when it was a first person pronoun. This finding provides converging support for the assumption that first person is a strong cue for actorhood, which can attenuate the behavioral reanalysis effect (for previous findings showing that strong cues for the target reading can attenuate reanalysis effects in behavioral data, though they are still observable electrophysiologically, see Bornkessel et al. 2004b).

Measures

The output from the implementation includes the selected distinctness metric as well as the scalar (weighted) difference in prominence for each item and condition (i.e. for each experimental stimulus). Additionally, the

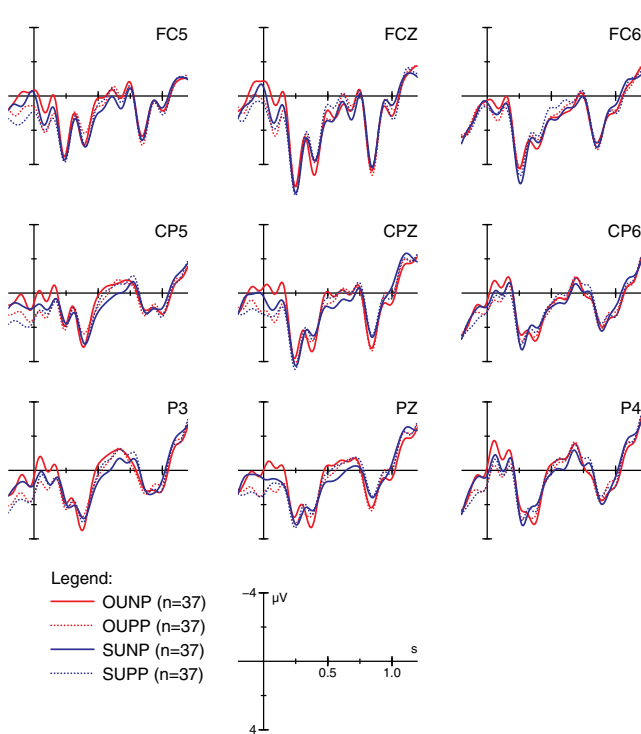


Fig. 4 Grand average ERPs triggered at the onset of NP2 for the unambiguous condition and NP2 a first person pronoun. The condition codes reflect the $2 \times 2 \times 2 \times 2$ design: S = subject (actor) initial word order, O = object (undergoer) initial; U = unambiguous, A = ambiguous; N = Noun, P = pronoun, for NP1 & NP2 respectively

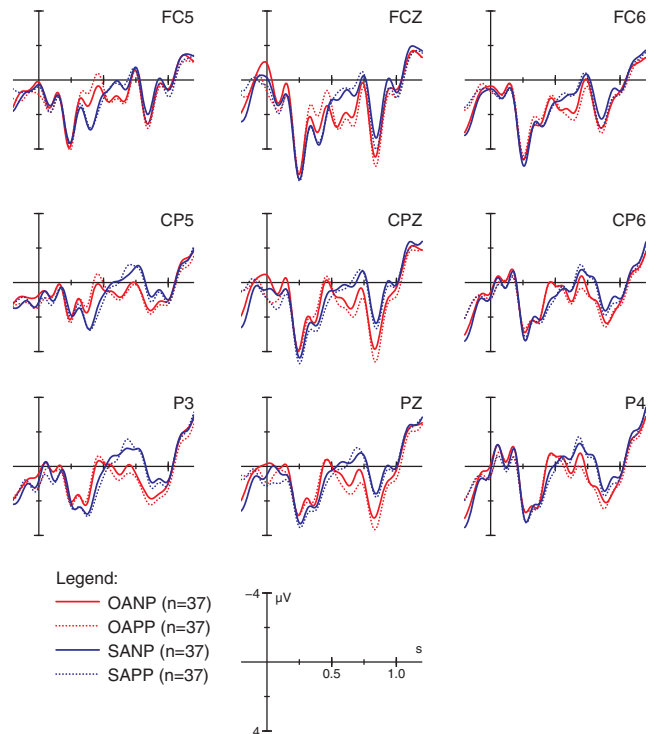


Fig. 5 Grand average ERPs triggered at the onset of NP2 for the ambiguous condition and NP2 a first person pronoun. The condition codes reflect the $2 \times 2 \times 2 \times 2$ design: S = subject (actor) initial word order, O = object (undergoer) initial; U = unambiguous, A = ambiguous; N = Noun, P = pronoun, for NP1 & NP2 respectively

implementation also outputs prominence scores calculated for NP1 and NP2. We duplicate this data for all subjects and enter it into the dataframe for EEG data. Based on visual inspection and significance testing via repeated measures analyses of variance (ANOVAs), we restricted the analysis to a subset of the data.

First, the time window for the N400 was found to be about 300–500ms post stimulus onset for the pronouns, and about 100ms later (400–600ms post onset) for the nouns (cf. Fig. 2 & 3 vs. Fig. 4 & 5). A similar effect was found for the late positivity (P600) time window: 600–800ms post onset for the pronouns and 700–900ms post onset for the nouns. As such, the time windows were parameterized in the model: “N400” vs “P600”, with the exact time window reflecting whether the target stimulus was a noun or a pronoun. This difference in latency is not unexpected: both frequency and length are known to influence the latency of exogenous ERP components. Pronouns being highly frequent and short (a classic example of Zipf’s Law; (Zipf 1935, 1949; Manning and Schütze 2000)) thus elicit a somewhat earlier effect. This is predicted by the cascaded architecture of the eADM: for shorter words, the information that is necessary for processing to proceed to the next step accrues more quickly.

Our choice of relatively traditional windows for the N400 and late positivity should thus not be taken as reflections of an a priori assumption about the ontological latency of these components. As previously mentioned, the decisive attribute of a component is its polarity; latency is to some extent an indication of the amount of processing necessary to reach the computational step reflected by a particular component. Amplitude is meaningful as a vague correlate of processing power needed at a particular step; however, due to well-known issues with equivalent dipoles, cancellation, etc., amplitude of scalp EEG is not a monotonic function of processing effort.

Furthermore, the ANOVA performed across five regions of interest (four quadrants and midline) revealed the strongest effects and interactions in the left posterior ROI, and for simplicity and computability, we restrict our model fitting to this ROI. The relevant ANOVA results are summarized in the Appendix (Tables 32 and 33).²⁰

²⁰More rigorous methods are available for dynamically determining the time window and topographical distribution of components. Maris (2004) and Maris and Oostenveld (2007) propose the necessary methods for non parametric method testing and determination of the effects in time and space (topography). Issues of computational tractability as well as data set size (different (sub)sets of data have to be used for determining the spatiotemporal distribution and testing it) reaffirmed our decision against introducing too many non-traditional methods for this initial computational model.

N400

We begin with the parametric time window “N400”. Here, we first generated the most basic models using only the distinctness measures as fixed effects. However, we note that neither `sdiff` nor `dist` explicitly encode the experimental parameter ambiguity, i.e. the degree of evidence for an actor or non-actor analysis of the first argument based on morphological case marking, the strongest cue to actorhood in German. In German, there are two possible ways of deterministically resolving locally ambiguous case marking: (a) the elimination of possibilities by another unambiguously marked argument and (b) agreement with the verb. It is, however, generally accepted in the psycholinguistic and neurolinguistic literature that sentence processing proceeds incrementally, i.e. the processing system uses strategies to resolve local ambiguities even in the absence of clear evidence for one or the other reading in the input (Marslen-Wilson 1973; Crocker 1994). In this experiment, ambiguity resolution was provided via (a) on the second argument, which means that the initial ambiguity affects both which predictions the language system is able to make initially and how much new information becomes available at NP2. It thus makes sense to see how the parameter ambiguity interacts with our distinctness measures. For `sdiff` (Table 4), a likelihood ratio test reveals a significant improvement (Table 5). Similarly, we find an improvement for `dist` (Table 6), albeit a smaller one (Table 7). The comparison between the models with ambiguity, which are shown in

Table 4 Summary of model fit for `sdiff` (weighted distinctness) and ambiguity in the N400 window

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
395191	395291	−197584	395169
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.18	0.42
	c.(sdiff)	2.3e−08	0.00015
subj	(Intercept)	1.2	1.1
	c.(sdiff)	2.6e−08	0.00016
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.2	0.19	6
ambiguityunambig	0.5	0.031	16
c.(sdiff)	−0.00039	4e−05	−9.8
ambiguityunambig:c.(sdiff)	0.00038	2.5e−05	15

Table 5 Statistics for models in the N400 time window based on the *sdiff* metric, showing the effect of ambiguity with interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<i>sdiff</i> : mean ~ c(<i>sdiff</i>) + ...	9	395675	-197828.97			
<i>sdiff.ambiguity</i> : mean ~ ambiguity * c(<i>sdiff</i>) + ...	11	395190	-197584.42	489.11	2	<2.2e-16 ***

(Random effect structure elided. See page 12)

Tables 4 and 6 for *sdiff* and *dist*, respectively, show that *sdiff* provides a better fit to the data.

Examining the models more closely, we see that the interaction between ambiguity and *dist* was not significant. In light of this missing interaction with *dist*, we can also consider using ambiguity as a simple model parameter that does not interact with our distinctness measures. In this case, we find a significant improvement for *dist* (Table 8) over the model without any ambiguity, and, moreover, the model with interaction does not differ significantly from the one without (Tables 9 and 10).

It thus appears that the interaction with ambiguity was particular to *sdiff*.

At this point, it is important to note that *sdiff* differs from *dist* not only in its weighting, but also in its use of “directionality” by being a signed value. We can also calculate a signed version of *dist*, termed *signdist*, by the sum of the pairwise differences: $\sum_i (NP2_i - NP1_i)$. This is the same as the Manhattan metric without absolute value signs or *sdiff* with all weights equal to one (via associativity of addition and subtraction, see

Footnote 13, p. 9). Intuitively, this measurement is the net change in prominence features—a negative value indicates fewer prominence features, while a positive value indicates more positive features. As with the other distance measures, the parameter ambiguity improves model fit significantly. Tellingly, the minimally adequate model for *signdist* with ambiguity (Table 11) does not differ from the minimally adequate model for *dist* (Table 12); see Table 13 for a direct comparison between the minimally adequate models for all 3 predictors.

Holding ambiguity constant to examine the interaction in more depth, we can again compare *sdiff* and *dist*. We find that they do not differ for unambiguous sentences; however, *sdiff* performs substantially better than *dist* and even *signdist* as a model predictor for ambiguous sentences (Table 14). This is immediately apparent in Figs. 6–10.

It is clear that *dist* behaves roughly the same, regardless of ambiguity, while *signdist* and *sdiff* interact with ambiguity—directionality clearly plays a role in the ambiguous condition. We even see that the slope in the ambiguous condition for *dist* is actually in the opposite direction of the other two predictors.

In Figs. 7 and 8, we can observe some difference between *signdist* and *sdiff*, in particular that the confidence interval is broader for *signdist*, which indicates a poorer fit. When we visualize the data in three dimensions as contour plots instead of as two subplots, the difference becomes even clearer (Figs. 9 and 10). Color indicates height, variation in color thus means variation in height, i.e. slope. Level curves, like in a topography map, indicate the overall shape of the landscape. The flat coloring in the unambiguous conditions for *signdist* and *sdiff* is indicative of the amount of variation being very small in comparison to the variation by the ambiguous condition. Moreover, *sdiff* shows a much more nuanced behavior in the ambiguous condition than *signdist*—this is clearly visible in the spacing between contour lines and their respective heights (difference between neighboring colors in the figures). The combination of weightedness and direction is much more telling about the processing of ambiguities than direction alone.

The *sdiff* model for the N400 time window reveals a strong negative correlation between *sdiff* and mean ERP

Table 6 Summary of model fit for *dist* (feature overlap) and ambiguity in the N400 window

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
395140	395240	-197559	395118
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.18	0.43
	c.(dist)	0.057	0.24
subj	(Intercept)	1.2	1.1
	c.(dist)	0.12	0.35
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.3	0.19	6.6
ambiguityunambig	0.26	0.036	7.2
c.(dist)	0.24	0.07	3.5
ambiguityunambig:c.(dist)	-0.012	0.036	-0.33

Table 7 Statistics for models in the N400 time window based on the `dist` metric, showing the effect of ambiguity with interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>dist: mean ~ c.(dist) + ...</code>	9	395188	-197585.02			
<code>dist.ambiguity: mean ~ ambiguity * c.(dist) + ...</code>	11	395139	-197558.87	52.31	2	4.37e-12 ***

(Random effect structure elided. See page 12)

amplitude, especially in the ambiguous condition (cf. sign of the *t*-statistic in Table 15, gradient direction in Fig. 8).

The decrease in the mean reflects the negativity in the ERP response, while the increase in prominence reflects an undergoer-first word order. For `signdist`, we see a weaker, yet similar effect. For `dist`, we see a positive correlation with ERP amplitude (`dist` is non negative per definition, but the models used centered values): the more features that don't overlap, the greater the mean, and hence, the smaller the negativity. In the unambiguous conditions, the correlation between mean ERP amplitude and `sdiff` is not significant (Table 16); however, we again see that the sign remains negative.

Additionally, the interaction of signedness and weightedness in `sdiff` expresses itself twofold. Signedness is a form of directionality and leads to a better gradient structure, and this becomes especially important when the correct directionality is not initially clear, namely in the ambiguous condition. This provides for the similarity in structure between `sdiff` and `signdist` that we see in Figs. 7 and 8 in contrast to the level structure of `dist` (Fig. 6). The weightedness of `sdiff` then contributes a

Table 8 Summary of model fit for `dist` (feature overlap) and ambiguity in the N400 window (without interaction)

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
395138	395229	-197559	395118
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.18	0.43
	<code>c.(dist)</code>	0.057	0.24
subj	(Intercept)	1.2	1.1
	<code>c.(dist)</code>	0.12	0.35
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.3	0.19	6.6
<code>ambiguityunambig</code>	0.26	0.036	7.2
<code>c.(dist)</code>	0.24	0.067	3.5

further, necessary granularity to the model. Directionality provides information about the direction of change, while weightedness contributes additional information about the amount of change.

From the models tested, the best is then the one using `sdiff` interacting with ambiguity for its fixed effects (Table 13).

Finally, in an additional post-hoc test, we can compare `sdiff` to a traditional, unweighted syntactic measure—i.e. a subject as opposed to an actor strategy—using new metrics, `syndist` and `syndist`, which are restricted to the features PERSON, NUMBER and CASE. The unweightedness follows from the all-or-nothing principles of agreement and case marking—either a verb and a noun agree or they don't / a noun is either nominative or it isn't. As is evident in Table 17, even under these experimental circumstances, without global ambiguity, where the deterministic case marking of German provides for a clear syntactic analysis, the prominence-based model fares better. Nevertheless, since cues to actorhood and subjecthood show considerable overlap in the present experimental design, this result can only be taken as a tentative initial indication that an actor-based strategy outperforms a subject-based strategy when the two are tested against each other via computational modelling.

Late Positivity (P600)

Visual inspection of the ERP data (Figs. 2–5) suggests a secondary effect in the form of a late positivity, which is line with previous findings on undergoer-initial ambiguous sentences in German (Haupt et al. 2008). As in the N400 time window, the models including ambiguity as an additional fixed factor perform far better than those without ambiguity. Similarly, `dist` shows no interaction with ambiguity (Tables 18 and 19), while `sdiff` and `signdist` do (Tables 20, 21, 22 and 23).

In accordance with the reverse in polarity over a biphasic reaction, we also see a reverse in effect direction: whereas a more highly positive `sdiff` correlated with a decreased mean (negativity) in the N400 time window, it correlates with an increased mean (positivity) in the P600 time window. Similarly, `dist` now correlates with a decreased mean (negativity, or here, lack of a positivity).

Table 9 Statistics for models in the N400 time window based on `dist` metric, showing the effect of ambiguity without interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>dist: mean ~ c.(dist) + ...</code>	9	395188	−197585.02			
<code>dist.ambiguity.no_int: mean ~ ambiguity + c.(dist) + ...</code>	10	395137	−197558.92	52.20	1	5.00e−13 ***

(Random effect structure elided. See page 12)

This is reflected in the respective *t*-statistics (Tables 18, 20 and 22): their signs have reversed.

A direct comparison of the minimally adequate models for each predictor can be found in Table 24.

Following the smaller effect size of the late positivity, the models do not differ by much. (The apparent trivial advantage for `dist` in AIC stems from it having fewer degrees of freedom, i.e. a smaller overfitting penalty.) However, upon resolving the interaction, we again see a greater differentiation in the ambiguous but not in the unambiguous condition²¹ (Tables 25–28). In the ambiguous conditions, `sdiff` outperforms `dist`—as in the N400 time window.

This is also clearly reflected in Figs. 11, 12 and 13 where the gradient reflected in the unambiguous condition differs from the unambiguous condition for `sdiff` but not `dist` (see also Figs. 14 and 15 for a comparison of `signdist` and `sdiff` split by ambiguity). Interestingly, the difference in variance in the mean is overall less in both conditions: this is reflected by the narrower confidence intervals (Figs. 11–13) and in the visibility of the color gradient for the unambiguous condition for `sdiff` (Fig. 15). The latter is indicative of the variance in the ambiguous condition being comparable enough to the unambiguous condition that the same scale provides the necessary resolution for both conditions. The reversal in effect direction is also apparent in the reversal of the color schemes for the contour plots.

While the effect in the P600 time window is smaller than in the N400, the general trend is nonetheless clear: increased prominence of the second argument compared to the first leads to an increase in the mean amplitude in the later time window. More succinctly, we see a positivity in the P600 time window for an object-initial word order. Taken together with the N400 for the object-initial word order, we have a biphasic pattern for object-initial sentences.

As discussed in section “The Extended Argument Dependency Model (eADM) and Actor-Centered Comprehension”, the eADM posits a functional distinction

²¹The slightly better performance of `dist` in this comparison of the unambiguous conditions is twofold: (1) it has fewer degrees of freedom and hence a smaller overfitting penalty in the AIC measure, and (2) the positive-only nature of `dist` lines up with the directionality of the positivity (but not the negativity).

between the two components comprising the biphasic pattern observed here. While the N400 is assumed to reflect actor competition per se (including its resolution), the late positivity is assumed to index the behavioral reorientation induced by subjectively significant (task-relevant) events. In the present study, sentences requiring a reanalysis of the actor-first preference entailed such a reorientation since the degree of actor competition was relevant for participants’ completion of the judgement task. As this explanation presupposes that (in contrast to the N400 effect) the positivity effect is reaction-locked rather than stimulus-locked, we computed an additional analysis in which we included logarithmically transformed mean reaction times per participant and condition into the mixed effects models as continuous predictors. While both models are greatly improved by including average reaction time by subject for each condition (i.e. single-subject averages) as a factor (Tables 29 and 30), the improvement is much greater (many orders of magnitude) for the late positivity window, as would be expected for an effect of task.

Discussion

We have presented a computational model that implements the actor strategy in language comprehension. The predictions of the model were tested against the results of an empirical study using event-related brain potentials (ERPs). Specifically, we examined the predictive capacity of two metrics for computing argument distinctness (i.e. degree of competition for the actor role): the unweighted distance measure `dist` (the Manhattan metric) and the weighted scalar difference measure `sdiff`. While both measures proved to be statistically significant predictors of N400 - late positivity amplitude, `sdiff` provided better model fits than `dist`. This was apparent particularly in ambiguous sentences, which, in some cases, called for a reanalysis towards an undergoer-initial order. Moreover, though this was not the primary focus of the present study, the current results provide an initial indication that the `sdiff` metric of actor computation provides a better fit to the electrophysiological data than a metric based purely on cues to syntactic subjecthood. They further show that the N400 and late positivity responses can be dissociated in that the latter is

Table 10 Statistics for models in the N400 time window based on `dist` metric, comparing the modelling of ambiguity with and without interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>dist.ambiguity.no_int: mean ~ ambiguity + c.(dist) + ...</code>	10	395137	−197558.92			
<code>dist.ambiguity: mean ~ ambiguity * c.(dist) + ...</code>	11	395139	−197558.87	0.11	1	7.41e−01

(Random effect structure elided. See page 12)

tied more closely to participants' behavioral reactions than the former. In the following, we will first discuss the evidence supporting a weighted as opposed to an unweighted distance metric and the architectural consequences arising from this result, before turning to implications for the functional interpretation of the N400 and late positivity in language processing tasks. We will then describe how this initial computational model of the actor strategy might serve to advance the development of a neurobiologically plausible model of actorhood computation. Finally, we will describe some future directions resulting from this work.

Evidence for and Consequences of a Weighted Distance Metric

As mentioned in section “[Distinctness/Actor Competition](#)”, `dist` (the Manhattan metric) roughly corresponds with “feature overlap” or traditional notions of similarity-based interference, such as those in memory-based models of

language processing (e.g. Lewis et al. 2006; Lewis 2000; Lewis and Vasishth 2005; McElree et al. 2003). By contrast, `sdiff` takes into account the language-specific weighting of the actor-related features. For German, this weighting places particular emphasis on the function of unambiguous case marking, which is the dominant cue to actor assignment when it is available (MacWhinney et al. 1984). In addition, it allows for coalitions of features to form to overcome “deficient” (ambiguous) case marking (Bates et al. 1982). The strong role of case as well as well the supplementary role of coalitions of weaker cues is most obvious in comparing the ambiguous and unambiguous conditions.

The predictive power of `sdiff` comes not only from its weighting, but also from its directionality, which serves to model the incremental demands of language processing, including the development (and possible fulfillment) of expectations. In relation to the current experiment, the directionality of the `sdiff` measure (negative or positive) essentially reflected the degree to which the parser's expectations about the prominence of the second argument were met. When `sdiff` was positive, the second argument was more prominent than the first, thus requiring a revision of the initial actor analysis of the first argument. This was reflected in a biphasic N400 - late positivity pattern, as was already observed in previous studies on actor-reanalysis in German (Haupt et al. 2008). As noted in section “[Evidence for the Actor Heuristic and for Competition for the Actor Role](#)” we interpret the negativity as an index of actor competition (leading to reanalysis of the initial actor-first preference in this case) and the late positivity as reflecting a behaviorally significant categorization of the sentences as less well-formed. This categorization reflects

Table 11 Summary of model fit for `signdist` (directed, net change in prominence features) and ambiguity in the N400 window

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
395396	395496	−197687	395374
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.17	0.42
	<code>c.(signdist)</code>	0.015	0.12
subj	(Intercept)	1.2	1.1
	<code>c.(signdist)</code>	0.015	0.12
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.2	0.19	6.1
<code>ambiguityunambig</code>	0.5	0.031	16
<code>c.(signdist)</code>	−0.2	0.03	−6.8
<code>ambiguityunambig: c.(signdist)</code>	0.2	0.019	11

Table 12 Statistics for the minimally adequate models for each unweighted predictor in the N400 time window

	Df	AIC	logLik
<code>dist.ambiguity.no_int: mean ~ ambiguity + c.(dist) + ...</code>	10	395137	−197558.92
<code>signdist.ambiguity: mean ~ ambiguity * c.(signdist) + ...</code>	11	395395	−197686.89

(Random effect structure elided. See page 12)

Table 13 Statistics for the minimally adequate models for each predictor in the N400 time window: *sdiff* and *signdist* interact with ambiguity, *dist* does not

	Df	AIC	logLik
<i>dist.ambiguity.no.int:</i> mean ~ ambiguity + c(<i>dist</i>) + ...	10	395137	-197558.92
<i>sdiff.ambiguity:</i> mean ~ ambiguity * c(<i>sdiff</i>) + ...	11	395190	-197584.42
<i>signdist.ambiguity:</i> mean ~ ambiguity * c(<i>signdist</i>) + ...	11	395395	-197686.89

(Random effect structure elided. See page 12)

the unmotivated positioning of the undergoer argument in a position that linearly precedes that of the actor.

Crucially, however, the advantage of *sdiff* over *dist* as a predictor of language-processing related neurophysiological activity cannot be reduced to the directionality of the *sdiff* metric. This was shown by the comparison of the two basic metrics with a directional (signed) version of *dist*, *signdist*. For locally ambiguous sentences, model fits involving *sdiff* were better than those involving both *dist* and *signdist*, thus attesting to the fact that both directionality and feature weighting contribute to the advantage for *sdiff*. Both N400 and late positivity amplitude for a revision of the actor-first preference were modulated by NP2 prominence (i.e. depending on whether NP2 was a first person pronoun or a definite noun phrase) and the magnitude of this modulation was predicted more accurately by the more fine-grained, weighted *sdiff* metric than by the unweighted *dist* and *signdist* metrics.

An additional divergence between the *sdiff* and *dist* measures is apparent in the model fits for the late positivity time window. Here, *sdiff* showed directionally opposite effects for ambiguous and unambiguous sentences: in the ambiguous cases, more positive *sdiff* correlated with higher positivity amplitude (as described above), while, for unambiguous sentences, more positive *sdiff* correlates with decreased positivity amplitude. By contrast, *dist* does not differentiate between ambiguous and unambiguous sentences, as demonstrated by the fact that the interaction of

Table 14 Statistics for models in the N400 time window with NP1 ambiguous

	Df	AIC	logLik
<i>dist:</i> mean ~ c(<i>dist</i>) + ...	9	197520	-98751.07
<i>sdiff:</i> mean ~ c(<i>sdiff</i>) + ...	9	197353	-98667.50
<i>signdist:</i> mean ~ c(<i>signdist</i>) + ...	9	197477	-98729.61

(Random effect structure elided. See page 12)

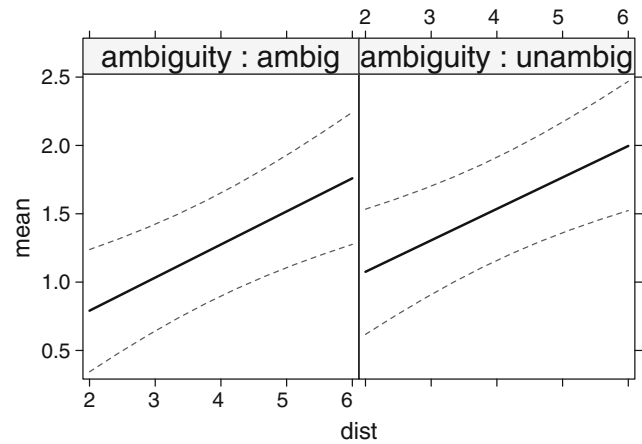


Fig. 6 Mean EEG in the N400 time window as a function of *dist* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95% confidence interval

dist and ambiguity can be removed from the model without affecting model fit. Thus, in unambiguous sentences, the data show a small, late positivity for actor- as opposed to undergoer-initial sentences. We posit that this could again be explained in terms of predictability in online processing. Specifically, unambiguous initial accusative marking (sentences with very strongly negative *sdiff*) allows for comparatively more prediction: in contrast to an initial nominative, it is apparent that the construction is transitive and that a second argument is required (Bornkessel et al. 2004a; Wolff et al. 2008). Accordingly, unambiguous nominative-initial (actor-initial) engender a slightly increased late positivity at the less predictable NP2 in comparison to their accusative-initial counterparts. Interestingly, we observed no such effect in the earlier time window. This supports the perspective that the N400 reflects competition for the

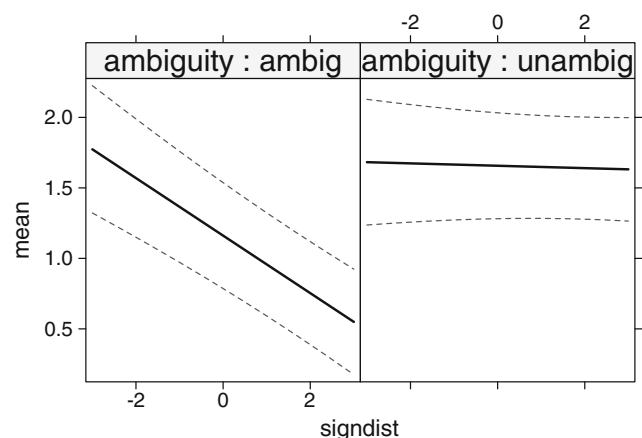


Fig. 7 Mean EEG in the N400 time window as a function of *signdist* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95% confidence interval

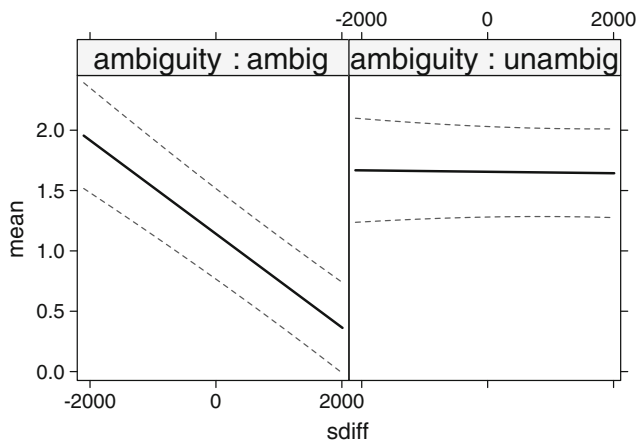


Fig. 8 Mean EEG in the N400 time window as a function of *sdiff* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95 % confidence interval

actor role more directly than the late positivity. In unambiguous sentences, competition is relatively low due to the strong weighting of unambiguous case marking information in German. Hence, no effects on N400 amplitude were observed in these sentence types. The late positivity, by contrast, reflects a behaviorally relevant well-formedness categorization, which can, in part, be envisaged as dependent on how predictable a particular element is within a given sentence context. This result emphasizes the qualitative difference between the N400 and late positivity effects, in spite of their tight interrelationship within the overall biphasic response. Such a difference is further supported by the finding that the amplitude of the late positivity effect showed a substantially stronger correlation with reaction

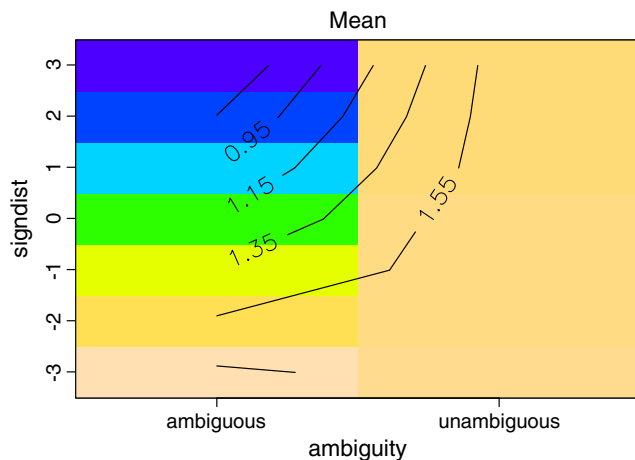


Fig. 9 Mean EEG in the N400 time window as a function of *signdist* (*centered*) and its interaction with ambiguity. The colors indicate the “height”, i.e., a range of (predicted) values of the mean EEG; the value is given by the contour curves. Colors that are closer together (e.g. light vs. dark blue) indicate finer differences. More color indicates more variation

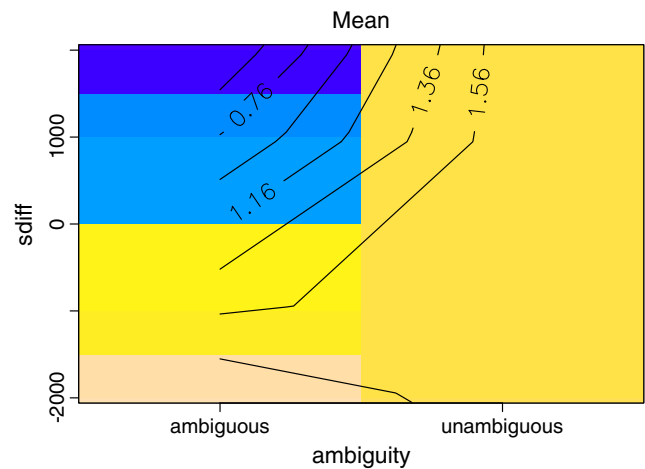


Fig. 10 Mean EEG in the N400 time window as a function of *sdiff* (*centered*) and its interaction with ambiguity. The colors indicate the “height”, i.e., a range of (predicted) values of the mean EEG; the value is given by the contour curves. Colors that are closer together (e.g. *light* vs. *dark blue*) indicate finer differences. More color indicates more variation

times for the behavioral task than the amplitude of the N400.

Overall, our findings suggest that the neural implementation of actor competition is best modeled by a weighted—rather than an unweighted—measure of the distance between the arguments in terms of actor features. This indicates that actor competition cannot be wholly reduced to similarity-based interference—at least in the sense of similarity-based interference as it is currently assumed in existing memory-based models of sentence processing. Crucially, similarity-based interference is a property of memory

Table 15 Summary of model fit for *sdiff* (weighted distinctness) in the N400 window, with NP1 ambiguous

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
197353	197429	−98668	197335
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.21	0.46
	c.(sdiff)	9.3e−08	0.0003
subj	(Intercept)	1.1	1.1
	c.(sdiff)	1.1e−07	0.00033
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.2	0.19	6.2
c.(sdiff)	−0.00039	7e−05	−5.6

Table 16 Summary of model fit for *sdiff* (weighted distinctness) in the N400 window, with NP1 unambiguous

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
197470	197546	−98726	197452
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.21	0.46
	c.(sdiff)	2.5e−08	0.00016
subj	(Intercept)	1.5	1.2
	c.(sdiff)	3.3e−08	0.00018
Residual		17	4.1
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	1.7	0.21	7.9
c.(sdiff)	−5.9e−06	3.8e−05	−0.16

models assuming a direct access to memory representations via a content-addressable pointer mechanism rather than a memory search: “The defining property of a content-addressable retrieval process is that information (cues) in the retrieval context enables direct access to relevant memory representations, without the need to search through extraneous memory representations” (McElree 2006, p. 163). Thus, since different types of cues serve to specify the “parts” making up the pointer address, they are not weighted—just as in a street address the name of the street, say, is not weighted differently to the house number or the post-code. It is therefore the qualitative overlap between cues that leads to similarity-based interference and weighting of the cues has no obvious role in a memory retrieval mechanism of this type. While, to the best of our knowledge, these characteristics apply to all existing models of language processing drawing on the assumption of direct memory access and similarity-based interference (McElree 2006; Lewis et al. 2006; Lewis and Vasishth 2005; Martin and McElree 2008), we cannot exclude that it may in principle

Table 17 Statistics comparing the predictors for a (unweighted) syntactic subject and (weighted) actor-prominence features in the N400 time window

	Df	AIC	logLik
synsigndist.ambiguity: mean ~ ambiguity * c.(synsigndist) + ...	11	395292	−197635.41
sdiff.ambiguity: mean ~ ambiguity * c.(sdiff) + ...	11	395190	−197584.42

(Random effect structure elided. See page 12)

Table 18 Summary of model fit for *dist* (feature overlap) and ambiguity in the P600 window

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
403101	403193	−201541	403081
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.066	0.26
	c.(dist)	0.068	0.26
subj	(Intercept)	1.2	1.1
	c.(dist)	0.15	0.39
Residual		19	4.4
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	0.47	0.18	2.6
ambiguityunambig	−0.093	0.038	−2.4
c.(dist)	−0.51	0.075	−6.9

be possible to assume a weighting of retrieval cues. If this were the case, the current findings could potentially also be subsumed under models based on the notion of similarity-based interference in memory retrieval. Nevertheless, we would like to stress once again that this assumption of feature weighting is not incorporated in any current models of this type.

Relation to Previous Work on Computer-Implemented Models

Vosse and Kempen (2008) conducted a similar computer-supported study using experimental stimuli from a previous experiment on scrambling (non canonical word order) effects with different verb classes in German, e.g. sentences where actorhood features do not as clearly align with syntactic subjecthood (Bornkessel and Schlesewsky 2006). The model presented is based primarily on morphosyntactic features—sentence topology (especially important for German data, but also used to model information structural alternations in word order), the related linear word order, and lexical features (including word category) and frames (related to valency in traditional linguistics). Input is processed via “unification” (see also Kempen, this issue), whereby trees are successively assembled and attached to one-another to build a single, unified tree / representation for the sentence. Processing difficulty is represented by cycles required to attach items in the correct configuration, e.g. a single nominative argument in agreement with the verb is readily attached, whereas a non agreeing noun-verb pair requires more effort to attach.

Table 19 Statistics for models in the P600 time window based on `dist` metric, comparing the modelling of ambiguity with and without interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>dist.ambiguity.no_int: mean ~ ambiguity + c.(dist) + ...</code>	10	403101	−201540.73			
<code>dist.ambiguity: mean ~ ambiguity * c.(dist) + ...</code>	11	403101	−201539.69	2.09	1	1.48e−01

(Random effect structure elided. See page 12)

In this way, Vosse and Kempen’s model is similar to the model presented here: both allow for a particular type of competition for attachment to a representation / role.²² The models differ however in which features are modeled as well as their ability to model the entire time-course. Crucially, Vosse and Kempen’s model primarily models late positivities (although they acknowledge that “certain negativities might find their origin in parser dynamics as well”) and fails to predict that subsequent studies have consistently shown a biphasic N400 - late positivity pattern following a reanalysis towards an undergoer-initial order in sentences with accusative verbs rather than only a positivity (Haupt et al. 2008).

However, it is important to point out that the aim of the present study was not to pit an actor-based interpretation strategy against a subject-centered interpretation strategy. Rather, based on the empirical motivation for an actor strategy in our own previous research (see section “[Evidence for the Actor Heuristic and for Competition for the Actor Role](#)”), it sought to examine the predictive capacity of various computational metrics designed to implement the actor heuristic. Thus, while the current findings provide an initial indication that a computational model based on an actor-centered rather than subject-centered interpretation strategy shows a superior fit to electrophysiological findings on human sentence comprehension (see the improvement of `sdiff` over `synsigndiff` in both time windows (Tables 17, and 31), the present experimental design included a considerable degree of overlap between the features relevant to the two strategies. Thus, a direct computational test of an actor-first strategy against a subject-first strategy in situations where the two diverge more strongly remains to be carried out in future research.

Implications for the Interpretation of the N400 and Late Positivity ERP Responses

The present findings have interesting and potentially important implications for the interpretation of language-related ERP responses. In this section, we will therefore relate our

²²Vosse and Kempen (2009) describe their parsing framework as a “dynamic model of syntactic parsing based on activation and inhibitory competition.”

results to current approaches to the N400 and late positivity in turn.

With regard to the N400, many researchers have recently come to favor a lexically-based interpretation of this component. According to this perspective, modulations of N400 amplitude do not reflect the computation of message-level meaning, but can rather be reduced to the effort required to retrieve a word from semantic memory (Kutas and Federmeier 2000). Effort is conditioned, in part, by intrinsic properties of the word such as its frequency, but also by its degree of preactivation given the preceding sentence and discourse context (Lau et al. 2008; Brouwer et al. 2012; Stroud and Phillips 2012). These assumptions can explain why N400 amplitude is modulated by single-word predictability (e.g. DeLong et al. 2005) and also why, in English and Dutch, “semantic reversal anomalies” (i.e. sentences such as “The hearty meals were devouring . . .”, Kim and Osterhout 2005) engender only late positivity effects but not N400 effects.

Table 20 Summary of model fit for `sdiff` (weighted distinctness) and ambiguity in the P600 window

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
403740	403841	−201859	403718
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.06	0.24
	c.(sdiff)	1.5e−08	0.00012
subj	(Intercept)	1.2	1.1
	c.(sdiff)	1.3e−08	0.00011
Residual		19	4.4
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	0.72	0.18	4
ambiguityunambig	−0.6	0.033	−18
c.(sdiff)	0.00053	3.4e−05	15
ambiguityunambig:c.(sdiff)	−0.00059	2.6e−05	−23

Table 21 Statistics for models in the P600 time window based on the `sdiff`, comparing the modelling of ambiguity with and without interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>sdiff.ambiguity.no.int: mean ~ ambiguity + c(sdiff) + ...</code>	10	404244	-202112.37			
<code>sdiff.ambiguity: mean ~ ambiguity * c(sdiff) + ...</code>	11	403740	-201859.03	506.68	1	< 2.2e-16 ***

(Random effect structure elided. See page 12)

Table 22 Summary of model fit for `signdist` (directed, net change in prominent features) and ambiguity in the P600 window

Linear mixed model fit by maximum likelihood

AIC	BIC	logLik	deviance
403875	403976	-201927	403853

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.059	0.24
	c.(signdist)	0.015	0.12
subj	(Intercept)	1.2	1.1
	c.(signdist)	0.0082	0.091
Residual		19	4.4

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	0.72	0.18	4
ambiguityunambig	-0.6	0.033	-18
c.(signdist)	0.31	0.028	11
ambiguityunambig:c.(signdist)	-0.36	0.02	-18

While this lexical view of the N400 is rather appealing and is able to account for a wide range of findings in the language-related ERP literature, it does not suffice to explain the present findings. Firstly, consider the basic finding of an increased N400 whenever a reanalysis towards an undergoer-initial order was required. This could be explained by the lexical view under the assumption that, following the actor interpretation of the first noun phrase and the subsequently encountered transitive verb, the processing system expects to encounter a second noun phrase marked for (or at least compatible with) accusative (rather

than nominative) case. In terms of preactivation, this would entail preactivating accusative case forms—either in terms of full-form lexical entries or of abstract, but nevertheless lexically stored, grammatical information. (But note that this explanation presupposes a rather specific view of lexical organization.) Crucially, however, it is not clear how this explanation might extend to the additional modulation of the actor-reanalysis effect via person / pronominality. The system has no way of predicting whether the second noun phrase will be a first person pronoun or a non-pronominal NP (since there is no expectation to encounter an actor argument at this point, one could not make the argument that a first person argument is more highly expected since it is a more prototypical instantiation of an actor argument). Thus, it is not clear how a purely lexically-based account of the N400 might account for the present findings (for further examples of problematic results for this class of N400 models, see Lotze et al. 2011; Bornkessel-Schlesewsky et al. 2011; Bourguignon et al. 2012). Rather, our data suggest that the N400—as one instance of a broader class of negativity responses—reflects at least certain aspects of integration between the current input and the input previously encountered. While top-down factors such as predictability, which can plausibly be translated into the notion of lexical preactivation, play an important part in determining N400 amplitude, bottom-up properties of the current input item must also be taken into account.

With regard to the late positivity, the close relationship between positivity amplitude and behavioral responses (reaction times) provides converging support for accounts of this component which posit a general (task-related) explanation rather than a specific linguistic function (e.g. reanalysis or effortful combinatorial analysis, Hagoort 2003; Kuperberg 2007). In addition to the account advocated here,

Table 23 Statistics for models in the P600 time window based on the `signdist`, comparing the modelling of ambiguity with and without interaction

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<code>signdist.ambiguity.no.int: mean ~ ambiguity + c(signdist) + ...</code>	10	404211	-202095.55			
<code>signdist.ambiguity: mean ~ ambiguity * c(signdist) + ...</code>	11	403875	-201926.73	337.66	1	< 2.2e-16 ***

(Random effect structure elided. See page 12)

Table 24 Statistics for the minimally adequate models for each predictor in the P600 time window: the models do not differ by much

	Df	AIC	logLik
dist.ambiguity.no_int: mean ~ ambiguity + c.(dist) + ...	10	403101	-201540.73
sdiff.ambiguity: mean ~ ambiguity * c.(sdiff) + ...	11	403740	-201859.03
signdist.ambiguity: mean ~ ambiguity * c.(signdist) + ...	11	403875	-201926.73

(Random effect structure elided. See page 12)

such a view has been proposed most prominently from the perspective of the conflict monitoring hypothesis (e.g. Kolk et al. 2003; van de Meerendonk et al. 2009). According to this proposal, late positivity effects in language processing reflect the detection of conflicting information and an ensuing check of the input for errors in previous processing steps. Evidence for this perspective stems, for example, from the finding that late positivities can be observed in response to various types of conflicts including orthographic errors (Vissers et al. 2006) and that, while both weak and strong semantic conflicts induce N400 effects, only strong conflicts engender an additional late positivity (van de Meerendonk et al. 2010). The conflict monitoring hypothesis can therefore also account for the observation that a reanalysis of the actor-first preference engenders late positivity effects (in addition to N400 modulations): here, conflict is high in comparison to sentences with an actor-initial word order. More precisely, in contrast to the N400, which is observable for all visually presented words (cf. the description in the very first study Kutas and Hillyard 1980), but can be described as an effect in certain contexts, the late positivity belongs more to the class of relative effects, occurring primarily in contrast to a condition with less (resolvable) conflict in experiments with a conflict-focused task (e.g. acceptability judgments) (Sassenhagen et al. 2013; Frenzel et al. 2011; Hahne and Friederici 2002).

However, a crucial difference between the conflict monitoring account and the present approach is that, according to the conflict monitoring view, the late positivity

Table 25 Statistics for models in the P600 time window with NP1 ambiguous

	Df	AIC	logLik
dist: mean ~ c.(dist) + ...	9	202377	-101179.75
sdiff: mean ~ c.(sdiff) + ...	9	202102	-101042.26
signdist: mean ~ c.(signdist) + ...	9	202396	-101189.33

(Random effect structure elided. See page 12)

Table 26 Statistics for models in the P600 time window with NP1 unambiguous

	Df	AIC	logLik
dist: mean ~ c.(dist) + ...	5	200678	-100334.17
sdiff: mean ~ c.(sdiff) + ...	5	200976	-100483.28
signdist: mean ~ c.(signdist) + ...	5	200981	-100485.95

(Random effect structure elided. See page 12)

Table 27 Summary of model fit for *sdiff* (weighted distinctness) in the P600 window, with NP1 ambiguous

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
202103	202179	-101042	202085
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.1	0.32
	c.(sdiff)	1.1e-07	0.00034
subj	(Intercept)	1.1	1
	c.(sdiff)	2.2e-07	0.00047
Residual		19	4.4
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	0.72	0.18	4
c.(sdiff)	0.00053	9.2e-05	5.7

Table 28 Summary of model fit for *sdiff* (weighted distinctness) in the P600 window, with NP1 unambiguous

Linear mixed model fit by maximum likelihood			
AIC	BIC	logLik	deviance
200977	201019	-100483	200967
Random effects:			
Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.1	0.32
subj	(Intercept)	1.4	1.2
Residual		19	4.3
Fixed effects:			
	Estimate	Std. Error	t value
(Intercept)	0.12	0.2	0.58
c.(sdiff)	-6.4e-05	1.2e-05	-5.6

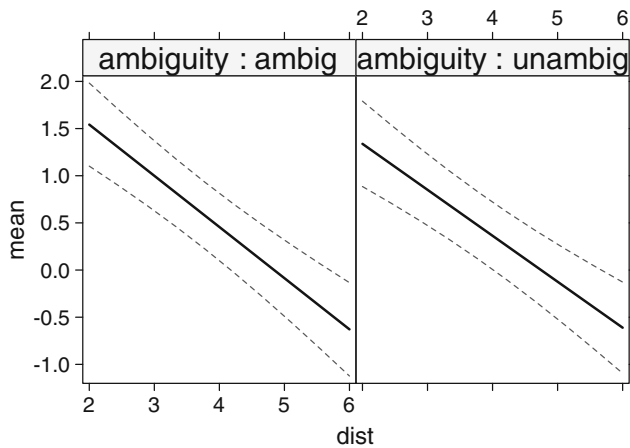


Fig. 11 Mean EEG in the P600 time window as a function of *dist* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95 % confidence interval

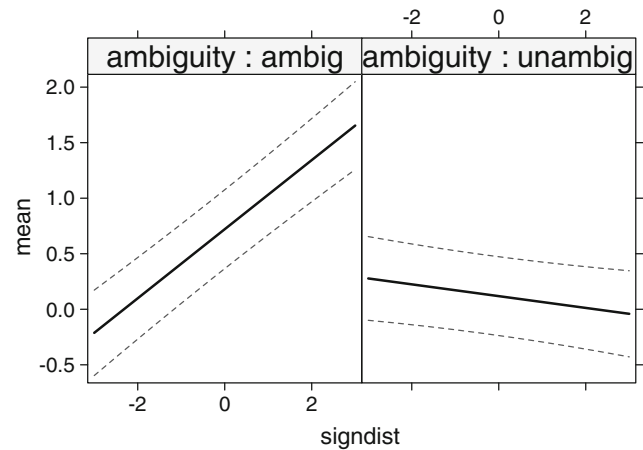


Fig. 12 Mean EEG in the P600 time window as a function of *signdist* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95 % confidence interval

reflects a (domain-general) reanalysis of the input (conflict resolution) (van de Meerendonk et al. 2009) while we posit that conflict resolution is reflected in the N400. Evidence for the N400 as the locus of conflict resolution stems from the observation that recent studies examining reanalyses of the actor-first preference have consistently found N400 effects, with additional late positivities depending on the behavioral relevance of the object-initial order (Bornkessel et al. 2004b; Haupt et al. 2008). Specifically, when the object-initial order was licensed by the presence of an object-experiencer verb and therefore did not call for a behavioral reorientation (i.e. judgement of the sentence as unacceptable), only an N400 effect was observed but no late positivity. The present results provide converging support for this perspective, since conflict resolution in the sense of a reanalysis should be more closely tied to the conflict-inducing feature in the input rather than to the behavioral response. Thus, the observation that the amplitude of the late positivity correlated considerably more strongly with the reaction times for the judgment task is expected under the assumption that the N400 reflects (input-related) conflict resolution, while the late positivity reflects the behavioral consequences of the conflict (and its resolution) in the given task environment.²³

²³Note that an explanation along these lines can also account for the dissociation between mild and strong conflicts observed by van de Meerendonk et al. (2010). As it appears plausible to assume that only the strong conflicts were registered as behaviorally significant, our account derives the finding of a late positivity for these conflicts, while no such effect was observed for mild conflicts. This explanation leads to the testable prediction that, with different task instructions (e.g. a judgment task emphasizing that even mild implausibilities should be classified as such), van de Meerendonk et al. (2010)’s mild conflict stimuli should also engender a late positivity.

Towards a Neurobiologically Realistic Computational Model of Actor computation

The present results demonstrate that the *sdiff* measure is a promising candidate for a neurocognitively plausible formalization of actor competition, as it is a valid predictor of neurophysiological activity related to sentence comprehension. In addition, we propose that this metric can be viewed as a first step towards a computational formalization of the neurobiological model described in section “[The Extended Argument Dependency Model \(eADM\) and Actor-Centered Comprehension](#)”. Specifically, we suggest that the insights gleaned from the present work may further our understanding of how linguistic categories posited within the eADM—such as the actor role—are recognized and processed in a neurobiologically plausible manner and how we might

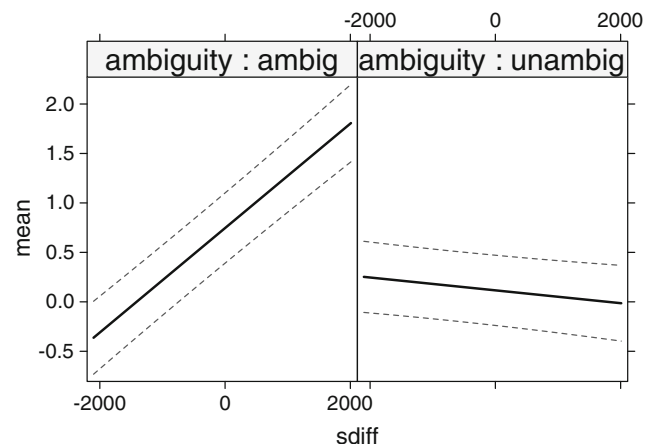


Fig. 13 Mean EEG in the P600 time window as a function of *sdiff* (*centered*) and its interaction with ambiguity. *Dashed lines* indicate the 95 % confidence interval

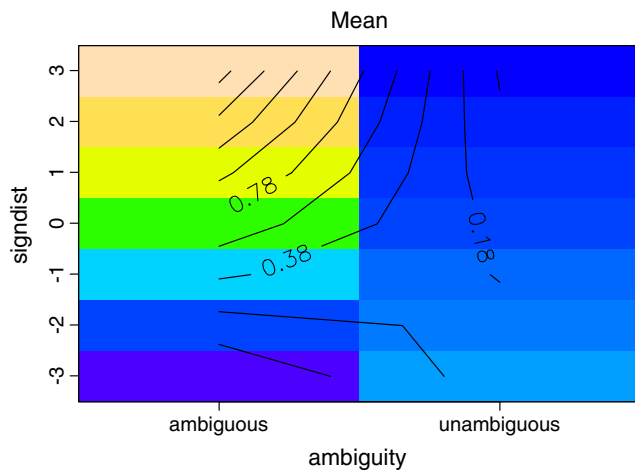


Fig. 14 Mean EEG in the P600 time window as a function of *signdist* (centered) and its interaction with ambiguity. The colors indicate the “height”, i.e., a range of (predicted) values of the mean EEG; the value is given by the contour curves. Colors that are closer together (e.g. light vs. dark blue) indicate finer differences. More color indicates more variation

envisage the relation between linguistic and non-linguistic categories.

Our proposal builds on the suggestion that, in view of its cross-linguistic ubiquity, actor could be modeled as an attractor category (Bornkessel-Schlesewsky and Schlewsky 2013a). Recent work in computational neuroscience has shown that attractor networks provide a neurobiologically plausible means of modeling decision-making processes (Deco et al. 2009; Deco et al. 2012), both for complex value-based choices as well as for perceptual classifications (“perceptual decision-making”; Heekeren et al.

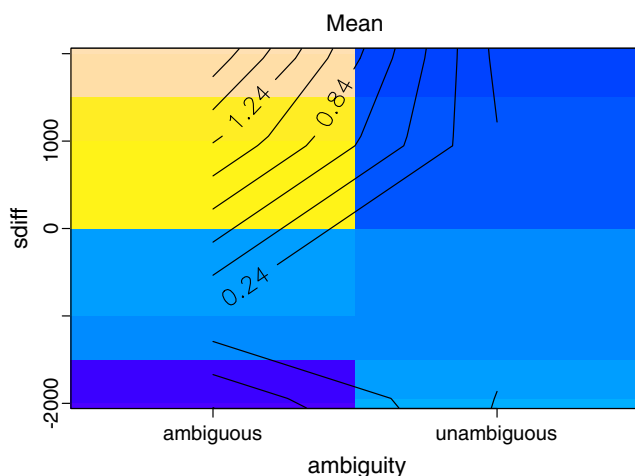


Fig. 15 Mean EEG in the P600 time window as a function of *sdiff* (centered) and its interaction with ambiguity. The colors indicate the “height”, i.e., a range of (predicted) values of the mean EEG; the value is given by the contour curves. Colors that are closer together (e.g. light vs. dark blue) indicate finer differences. More color indicates more variation

2004; Basten et al. 2010). In an attractor network, decisions can be modeled via attractor states in a neural network which are associated with (stable) high firing rates. Which state “wins” during decision making is determined by the current input and the initial stochastic firing behaviour of the network. Based on the overarching (language-independent) importance of the actor category (see section “Evidence for the Actor Heuristic and for Competition for the Actor Role”), it has been proposed that an attractor network for actor categorization exists independently of language (Bornkessel-Schlesewsky and Schlewsky 2013a). This network is universal, as it reflects the general human ability to recognize goal-directed action. The stable firing patterns inherent to this network are plausibly based on sets of input features that co-occur in domain-general actor recognition. As the linguistic actor category overlaps to a certain degree with these general features (e.g. via the features +HUMAN, +ANIMATE and +1ST. PERSON), there is a propensity for actor recognition via the general attractor network. With regard to more language-specific features (e.g. case marking), the system will learn that these correlate with the remaining (domain-general) actor features such that, in the mature system, they also push the network towards the actor recognition attractor state.²⁴ Crucially, an important consideration pertains to the degree of evidence for a certain decision—e.g. the classification of a certain event participant as an actor—that the current input offers. The *sdiff* metric can be viewed as a measure that captures this degree of evidence.

Weighted, directed measures, such as *sdiff*, provide the means to quantify the effects of attractor basins. Indeed, the physical metaphor behind attractor basins also provides insight into why *sdiff* functions better than *dist* or *signdist* (Fig. 16). The proximity of an attractor is given by *dist*, but not whether it is a positive or negative attractor (true attractor vs. repulsor, or hill vs. basin; see Fig. 17). This is a decent first approximation, but quickly fails in more rolling landscapes, e.g. in languages with free word order, where it is not clear which argument will come first. The directionality of *signdist* provides a better approximation, modeling attraction and repulsion, but the best approximation comes from the strength of the attractor (the steepness of the sides of the basin / hill or equivalently, the height and depth; see Fig. 18). This is exactly what *sdiff* does—the weightedness distorts the actor space, creating stronger and weaker attractors (Fig. 19). In this sense, deterministic case marking and garden path sentences are examples of attractor basins that are too deep to escape,

²⁴This proposal of a tight interrelationship between domain-general and linguistic actor features is supported by the recent observation that properties of an ideal actor may depend—at least to some degree—on the characteristics of one’s native language (Fausey et al. 2010; Fausey and Boroditsky 2011).

Table 29 Statistics for the minimally adequate (stimulus-based) models in the N400 time window compared to their extension via reaction time (RT)

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
sdiff.ambiguity: mean ~ ambiguity * c.(sdiff) + ...	11	395190	−197584.42			
sdiff.ambiguity.rt: mean ~ ambiguity * c.(sdiff) + log(rt) + ...	12	395172	−197574.02	20.80	1	<5.10e−06 ***

(Random effect structure elided. See page 12)

where the language system becomes trapped at the bottom of a well, or perhaps, to use another meaning of the word “space”, in a black hole.

Finally, though we have focused on the actor role here, we propose that the notion of attractor basins could be used to formalize the entire processing architecture shown in Fig. 1. Specifically, attractors could be used to model the categories assumed at every processing step within the cascade (e.g. phonemes, actor-event schemata etc.). They could further help to address an issue that is conspicuously missing from the current model implementation, namely the need to provide an estimate of the timing of the different processing steps and, accordingly, of the neurophysiological responses elicited by them. At present, the model only specifies the relative order of information processing but offers no quantifiable timing estimates. However, combining the assumption of cascaded, hierarchically organized processing steps and the attractor notion opens up a possible avenue for such a quantification. As noted in section “[The Extended Argument Dependency Model \(eADM\) and Actor-Centered Comprehension](#)”, cascaded processing is based on the idea that, once a sufficient degree of information has accrued, processing can proceed to the next step. Drawing upon the attractor notion, we can posit that the faster the system recognizes that information is relevant for a particular attractor, the faster processing at the step relevant to that attractor will be. Accordingly, the formalization of actor space presented here could be used as the basis for estimating processing latency as well as amplitude and, in our view, should also carry over to other linguistic categories. Of course, timing estimates will not be trivial given the different levels of neuronal responses that need to be considered here: as mentioned in section “[The Extended Argument Dependency Model \(eADM\) and Actor-Centered Comprehension](#)”, scalp

ERPs as examined here are macroscopic responses with (typically) multiple underlying sources and therefore cannot be directly compared to the cascade of activity that is assumed to proceed through individual regions along the antero-ventral and postero-dorsal streams. Accordingly, latencies of language-related ERP components such as the N400 likely do not reflect the absolute timing of information processing (see Bornkessel-Schlesewsky and Schlewsky 2013b). Nevertheless, assuming that our proposal regarding the basic relationship between evidence for an attractor and duration of the processing step in question is correct, both the direct neuronal responses and the neurophysiological responses measured by means of scalp EEG recordings should be quantifiable as some function of the degree of evidence for the respective attractor category.

Future Directions and the Role of Neuroinformatics

In the experiment presented here, `sdiff` showed the advantage of a weighted, directed distinctness measure over simple (unweighted) interference measures. Nonetheless, morphological case and ambiguities involving the same dominated the most important prominence variations. In future work, we aspire to test the metric against a wider range of stimuli, including globally ambiguous sentences and generally more naturalistic language. On account of the modular nature of the implementation, any EEG dataset could be processed and analyzed, either via adapting / constructing a suitable front-end (Stage 1) parser or via manually tagging the stimuli appropriately for Stage 2. As more work is done in this direction of quantifying linguistic differences in the brain, it becomes increasingly important to have diverse test data, especially if learning is to be

Table 30 Statistics for the minimally adequate (stimulus-based) models in the P600 time window compared to their extension via reaction time (RT)

	Df	AIC	logLik	Chisq	Chi Df	Pr(>Chisq)
sdiff.ambiguity: mean ~ ambiguity * c.(sdiff) + ...	11	403740	−201859.03			
sdiff.ambiguity.rt: mean ~ ambiguity * c.(sdiff) + log(rt) + ...	12	403505	−201740.85	236.36	1	<2.2e−16 ***

(Random effect structure elided. See page 12)

Table 31 Statistics comparing the predictors for a (unweighted) syntactic subject and (weighted) actor-prominence features P600 time window

	Df	AIC	logLik
synsigndist.ambiguity: mean \sim ambiguity * c.(synsigndist) + ...	11	404065	-202021.65
sdiff.ambiguity: mean \sim ambiguity * c.(sdiff) + ...	11	403740	-201859.03

(Random effect structure elided. See page 12)

simulated at some point. To this end, it is crucially important that databases of EEG data for diverse stimuli from typologically varied languages are available, similar to the corpora and treebanks used by researchers in natural language processing. A general model of language comprehension is the goal, not a model for particular dataset.

Call for Data (Banks)

To this end, we would like to see databanks of neuroanatomical and neurophysiological data similar to the “treebanks” common in computational linguistics and natural language processing research. Such databanks should provide a validated, state-of-the-art analysis with traditional methodologies, e.g. ANOVA (including standardized ROIs and/or single electrodes as a factor, with grand-average ERP) as well as parametric labeling of relevant linguistic information—cloze probability of each word, morphosyntactic features, thematic relations, known important semantic features (e.g. animacy and ideally other features that are expressed morphosyntactically in any of the world’s languages), lexical

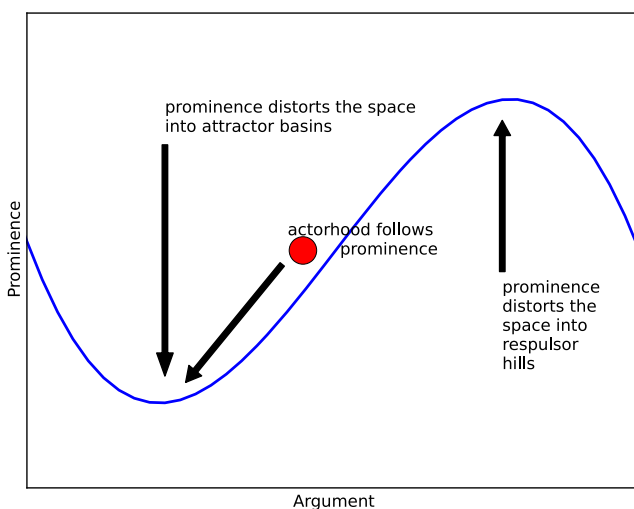


Fig. 16 Attractor basins in actor space. Prominence can be viewed as a distortion of actor space. The curvature of actor space then pulls or pushes actorhood towards a particular argument

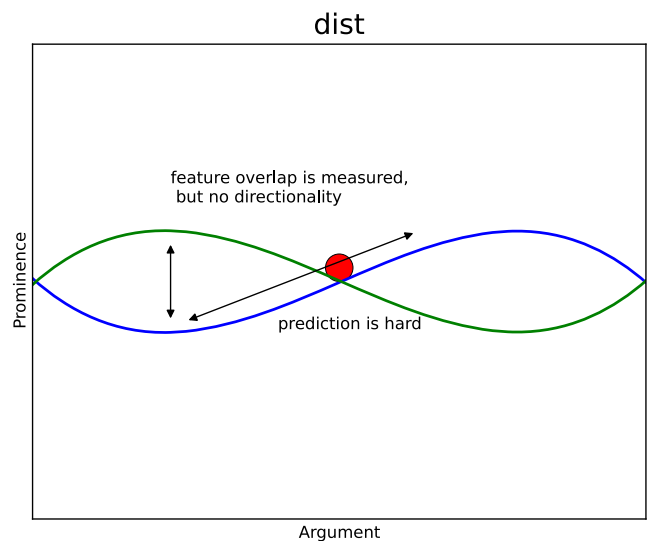


Fig. 17 Attractor basins in actor space as measured by *dist*. The directionality of distortion is lost, making prediction difficult

frequency estimates, estimates of syntactic frequency for any syntactic peculiarities (e.g. non canonical word orders), etc. The EEG data should preferably be stored in an open format, or at least in a format for which there are suitable plugins and converters—perhaps one of the formats supported by the open source EEGLAB software package. Data should not be filtered, rereferenced or otherwise manipulated offline before storage so as to not limit analysis by alternative techniques (time-frequency analysis, ICA, etc.). Instead, the measurement parameters (sampling rate, reference electrode, equipment manufacturer), experimental setup (presentation mode and aspects pertaining thereto) and anonymized subject data (age, sex, etc.) as well as

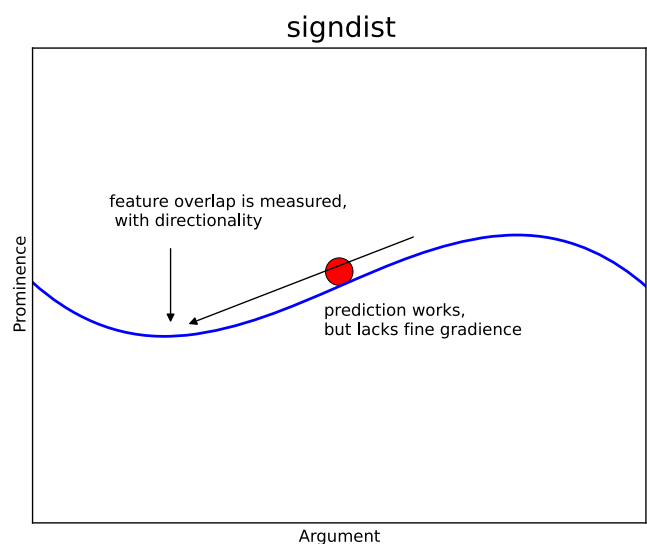


Fig. 18 Attractor basins in actor space as measured by *signdist*. The gradience of distortion is lost, leading to only qualitative predictions

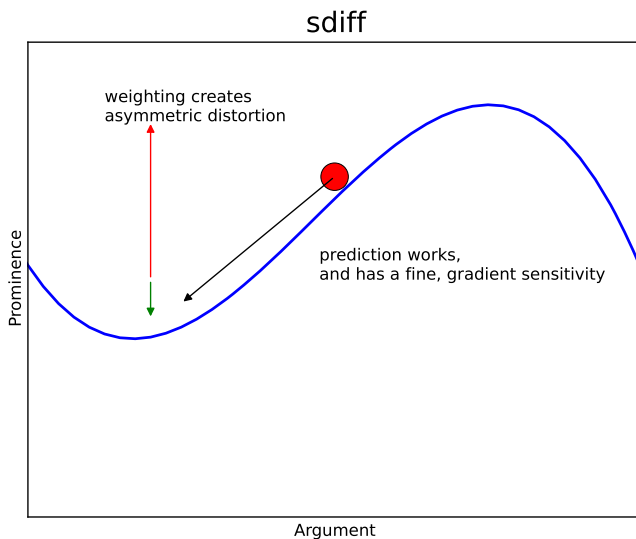


Fig. 19 Attractor basins in actor space as measured by *sdiff*. Both directionality and gradient of the distortion is preserved, thus enabling accurate prediction. The distortion here is to scale for the sentence *Die Bettlerin bedrängte den Kommissar auf der Straße*. (See Table 3)

experimental task, handedness of the test subject and task interface layout, should be stored as metadata. Optionally, the number of channels could be stored as well; however, this is not necessary. Channel names should be standardized to the 10-10/10-20 system terminology. Behavioral data should be linked not just as metadata, but also as a proper dataset unto itself. (Relational databases provide a convenient way to do this.) Only then, can we truly test our models of sentence comprehension, i.e. our “parsers”, with the same rigor that has been standard in other computational disciplines for years now—with lots of large, standardized tests.

Vosse and Kempen (2008) took an important first step in this direction, analyzing data from another experiment; however, it appears important to go beyond comparisons of modeling output with the published, summarized data. As discussed by Arbib et al. (this issue), it is important to remember that summary data implies the existence of non summarized data, i.e. more complete data. As one researcher’s noise is another’s signal, even the most basic filtering removes important data; the usual presentation of means and ANOVA leads even more to be desired. The BrainMap database is an excellent start for fMRI data, but it only makes the want of a comparable database for EEG data more striking. Recent trends in Open Access and pre-registration point to such databases as being the way of the future. We need data sharing beyond rebuttal and as common proving ground beyond the current experiments hand crafted to show off a particular model feature.

Beyond the traditional, well parameterized experimental data, we would ideally also like to see a complementary set of data acquired in a less structured, free-task environment.

That is, we would like to see a similar dataset of EEG/fMRI recording of natural stories with a maximal task of a few comprehension questions taken offline after the main experiment (Whitney et al. 2009) but with fully tagged input. Computational linguists use more than sets of simple, constructed sentences to test their data and so should we—our models need to be able to handle the full complexity of human language in its actual use and not just in our idealized laboratory conditions. These more complex inputs also present us the chance to move beyond sentence-processing models towards language-processing models.

The existence of large, standardized datasets also provides for a proving ground for newer methodologies. For example, although time-frequency analysis, principal component analysis (PCA) and independent component analysis (ICA) have been used in recent years to differentiate certain subtleties not readily apparent from traditional ERP-based analyses, the world of EEG-data is still dominated by ERP. This is almost certainly related to not just the complexity of these new methods, but also their unclear relationship to ERP results. A standard dataset provides exactly the playground necessary to demonstrate and test new methodologies and their relationship to old ones.

Brief Technical Notes on Implementation

The present implementation is in Python 3.2. A previous version was written and tested in Python 2.7; however, the implementation of Stage 1 and necessity of using non ASCII encoding for German sentence data motivated the shift to the 3.x series of Python with its much more extensive Unicode support. File and directory manipulations were all tested on POSIX compatible platforms.

There are options to set the baseline weights all equal to one (no weighting) or to a priori estimated weights based on previous work done in German (Kempe and MacWhinney 1999). A further correction (from empirical data) can then be applied to the individual baseline weights via additional options.

To test the weight configuration, a set of potential constructions in German is provided in a form directly processable by Stage 2. A test mode operating purely on these preanalyzed inputs is one of three modes of operation.

The other modes are a batch mode for generating predictions about experimental stimuli and an interactive mode for demonstrations of the model, as well as a mode capable of processing Stage 1 output vectors, either as list of experimental conditions from a file or interactively. Both the batch and interactive modes use a limited version of Stage 1, featuring a small parser customized for the experiment in question (see Stage 1).

More detailed documentation can be found with the publicly available source code (see below).

as the necessary input files is available to the public at <https://bitbucket.org/palday/ginnungagap-code/>.

Information Sharing Statement

All program source code for the implementation and generation of figures related to the mixed-models as well

Acknowledgments We would like to thank Rick Lewis and Joakim Nivre for valuable discussions and suggestions related to the development of the computational model. We would also like to thank Isabel Plauth for the data acquisition.

Appendix

Table 32 ANOVA for the N400 window

ANOVA:							
	Effect	DFn	DFd	F	p	p<.05	ges
2	roi	4.00	144.00	14.47	0.00	*	0.03
3	wordOrder	1.00	36.00	19.60	0.00	*	0.01
4	ambiguity	1.00	36.00	19.11	0.00	*	0.01
5	np1type	1.00	36.00	5.61	0.02	*	0.00
6	np2type	1.00	36.00	54.44	0.00	*	0.14
7	roi:wordOrder	4.00	144.00	5.51	0.00	*	0.00
8	roi:ambiguity	4.00	144.00	7.51	0.00	*	0.00
9	wordOrder:ambiguity	1.00	36.00	21.84	0.00	*	0.01
10	roi:np1type	4.00	144.00	13.25	0.00	*	0.00
11	wordOrder:np1type	1.00	36.00	0.57	0.45		0.00
12	ambiguity:np1type	1.00	36.00	0.75	0.39		0.00
13	roi:np2type	4.00	144.00	55.78	0.00	*	0.06
14	wordOrder:np2type	1.00	36.00	0.35	0.56		0.00
15	ambiguity:np2type	1.00	36.00	0.17	0.68		0.00
16	np1type:np2type	1.00	36.00	0.17	0.68		0.00
17	roi:wordOrder:ambiguity	4.00	144.00	2.58	0.04	*	0.00
18	roi:wordOrder:np1type	4.00	144.00	1.04	0.39		0.00
19	roi:ambiguity:np1type	4.00	144.00	0.53	0.72		0.00
20	wordOrder:ambiguity:np1type	1.00	36.00	1.30	0.26		0.00
21	roi:wordOrder:np2type	4.00	144.00	12.10	0.00	*	0.00
22	roi:ambiguity:np2type	4.00	144.00	4.53	0.00	*	0.00
23	wordOrder:ambiguity:np2type	1.00	36.00	6.40	0.02	*	0.00
24	roi:np1type:np2type	4.00	144.00	1.21	0.31		0.00
25	wordOrder:np1type:np2type	1.00	36.00	1.99	0.17		0.00
26	ambiguity:np1type:np2type	1.00	36.00	0.23	0.63		0.00
27	roi:wordOrder:ambiguity:np1type	4.00	144.00	0.61	0.66		0.00
28	roi:wordOrder:ambiguity:np2type	4.00	144.00	2.68	0.03	*	0.00
29	roi:wordOrder:np1type:np2type	4.00	144.00	0.35	0.84		0.00
30	roi:ambiguity:np1type:np2type	4.00	144.00	4.03	0.00	*	0.00
31	wordOrder:ambiguity:np1type:np2type	1.00	36.00	2.88	0.10		0.00
32	roi:wordOrder:ambiguity:np1type:np2type	4.00	144.00	0.91	0.46		0.00

Table 32 (continued)**Sphericity Corrections:**

	Effect	W	p	p<.05
2	roi	0.37	0.00	*
7	roi:wordOrder	0.20	0.00	*
8	roi:ambiguity	0.28	0.00	*
10	roi:np1type	0.22	0.00	*
13	roi:np2type	0.33	0.00	*
17	roi:wordOrder:ambiguity	0.22	0.00	*
18	roi:wordOrder:np1type	0.32	0.00	*
19	roi:ambiguity:np1type	0.07	0.00	*
21	roi:wordOrder:np2type	0.19	0.00	*
22	roi:ambiguity:np2type	0.23	0.00	*
24	roi:np1type:np2type	0.12	0.00	*
27	roi:wordOrder:ambiguity:np1type	0.22	0.00	*
28	roi:wordOrder:ambiguity:np2type	0.06	0.00	*
29	roi:wordOrder:np1type:np2type	0.14	0.00	*
30	roi:ambiguity:np1type:np2type	0.32	0.00	*
32	roi:wordOrder:ambiguity:np1type:np2type	0.12	0.00	*

Table 33 ANOVA for the N400 time window resolved in the Left-Posterior Region of Interest**ANOVA:**

	Effect	DFn	DFd	F	p	p<.05	ges
2	roi	4.00	144.00	24.10	0.00	*	0.06
3	wordOrder	1.00	36.00	25.71	0.00	*	0.02
4	ambiguity	1.00	36.00	31.79	0.00	*	0.04
5	np1type	1.00	36.00	0.16	0.69		0.00
6	np2type	1.00	36.00	93.33	0.00	*	0.19
7	roi:wordOrder	4.00	144.00	7.46	0.00	*	0.00
8	roi:ambiguity	4.00	144.00	5.01	0.00	*	0.00
9	wordOrder:ambiguity	1.00	36.00	37.31	0.00	*	0.04
10	roi:np1type	4.00	144.00	0.27	0.89		0.00
11	wordOrder:np1type	1.00	36.00	0.11	0.74		0.00
12	ambiguity:np1type	1.00	36.00	2.62	0.11		0.00
13	roi:np2type	4.00	144.00	3.44	0.01	*	0.00
14	wordOrder:np2type	1.00	36.00	0.87	0.36		0.00
15	ambiguity:np2type	1.00	36.00	1.98	0.17		0.00
16	np1type:np2type	1.00	36.00	1.42	0.24		0.00
17	roi:wordOrder:ambiguity	4.00	144.00	9.99	0.00	*	0.00
18	roi:wordOrder:np1type	4.00	144.00	0.74	0.56		0.00
19	roi:ambiguity:np1type	4.00	144.00	2.76	0.03	*	0.00
20	wordOrder:ambiguity:np1type	1.00	36.00	0.23	0.63		0.00
21	roi:wordOrder:np2type	4.00	144.00	0.72	0.58		0.00
22	roi:ambiguity:np2type	4.00	144.00	1.82	0.13		0.00
23	wordOrder:ambiguity:np2type	1.00	36.00	0.50	0.48		0.00
24	roi:np1type:np2type	4.00	144.00	0.20	0.94		0.00

Table 33 (continued)

	Effect	DFn	DFd	F	p	p<.05	ges
25	wordOrder:np1type:np2type	1.00	36.00	5.24	0.03	*	0.00
26	ambiguity:np1type:np2type	1.00	36.00	0.86	0.36		0.00
27	roi:wordOrder:ambiguity:np1type	4.00	144.00	1.71	0.15		0.00
28	roi:wordOrder:ambiguity:np2type	4.00	144.00	1.85	0.12		0.00
29	roi:wordOrder:np1type:np2type	4.00	144.00	0.53	0.72		0.00
30	roi:ambiguity:np1type:np2type	4.00	144.00	0.07	0.99		0.00
31	wordOrder:ambiguity:np1type:np2type	1.00	36.00	0.89	0.35		0.00
32	roi:wordOrder:ambiguity:np1type:np2type	4.00	144.00	1.89	0.11		0.00

Sphericity Corrections:

	Effect	W	p	p<.05
2	roi	0.48	0.00	*
7	roi:wordOrder	0.20	0.00	*
8	roi:ambiguity	0.24	0.00	*
10	roi:np1type	0.15	0.00	*
13	roi:np2type	0.53	0.01	*
17	roi:wordOrder:ambiguity	0.34	0.00	*
18	roi:wordOrder:np1type	0.23	0.00	*
19	roi:ambiguity:np1type	0.13	0.00	*
21	roi:wordOrder:np2type	0.17	0.00	*
22	roi:ambiguity:np2type	0.32	0.00	*
24	roi:np1type:np2type	0.20	0.00	*
27	roi:wordOrder:ambiguity:np1type	0.21	0.00	*
28	roi:wordOrder:ambiguity:np2type	0.41	0.00	*
29	roi:wordOrder:np1type:np2type	0.12	0.00	*
30	roi:ambiguity:np1type:np2type	0.15	0.00	*
32	roi:wordOrder:ambiguity:np1type:np2type	0.14	0.00	*

Table 34 ANOVA for the P600 window

ANOVA:							
	Effect	DFn	DFd	F	p	p<.05	ges
2	wordOrder	1.00	36.00	23.01	0.00	*	0.02
3	ambiguity	1.00	36.00	30.55	0.00	*	0.03
4	np1type	1.00	36.00	0.01	0.93		0.00
5	np2type	1.00	36.00	9.65	0.00	*	0.04
6	wordOrder:ambiguity	1.00	36.00	20.21	0.00	*	0.02
7	wordOrder:np1type	1.00	36.00	1.96	0.17		0.00
8	ambiguity:np1type	1.00	36.00	0.23	0.64		0.00
9	wordOrder:np2type	1.00	36.00	4.41	0.04	*	0.00
10	ambiguity:np2type	1.00	36.00	0.01	0.93		0.00
11	np1type:np2type	1.00	36.00	1.31	0.26		0.00
12	wordOrder:ambiguity:np1type	1.00	36.00	0.50	0.48		0.00
13	wordOrder:ambiguity:np2type	1.00	36.00	5.17	0.03	*	0.00
14	wordOrder:np1type:np2type	1.00	36.00	2.01	0.17		0.00
15	ambiguity:np1type:np2type	1.00	36.00	0.03	0.86		0.00
16	wordOrder:ambiguity:np1type:np2type	1.00	36.00	3.00	0.09		0.00

Table 35 ANOVA for the P600 time window resolved in the Left-Posterior Region of Interest

ANOVA:							
	Effect	DFn	DFd	F	p	p<.05	ges
2	wordOrder	1.00	36.00	28.20	0.00	*	0.02
3	ambiguity	1.00	36.00	26.00	0.00	*	0.04
4	np1type	1.00	36.00	0.00	1.00		0.00
5	np2type	1.00	36.00	85.25	0.00	*	0.23
6	wordOrder:ambiguity	1.00	36.00	43.74	0.00	*	0.05
7	wordOrder:np1type	1.00	36.00	0.56	0.46		0.00
8	ambiguity:np1type	1.00	36.00	1.33	0.26		0.00
9	wordOrder:np2type	1.00	36.00	2.09	0.16		0.00
10	ambiguity:np2type	1.00	36.00	2.88	0.10		0.00
11	np1type:np2type	1.00	36.00	1.30	0.26		0.00
12	wordOrder:ambiguity:np1type	1.00	36.00	0.50	0.48		0.00
13	wordOrder:ambiguity:np2type	1.00	36.00	0.15	0.70		0.00
14	wordOrder:np1type:np2type	1.00	36.00	7.45	0.01	*	0.00
15	ambiguity:np1type:np2type	1.00	36.00	0.85	0.36		0.00
16	wordOrder:ambiguity:np1type:np2type	1.00	36.00	0.00	0.95		0.00

Table 36 Summary statistics for the accuracy in trials

Summary statistics:							
wordOrder	ambiguity	np1type	np2type	N	Mean	SD	FLSD
O	A	N	N	37.00	0.86	0.13	0.01
O	A	N	P	37.00	0.93	0.08	0.01
O	A	P	N	37.00	0.89	0.11	0.01
O	A	P	P	37.00	0.94	0.06	0.01
O	U	N	N	37.00	0.95	0.06	0.01
O	U	N	P	37.00	0.94	0.04	0.01
O	U	P	N	37.00	0.98	0.02	0.01
O	U	P	P	37.00	0.96	0.02	0.01
S	A	N	N	37.00	0.97	0.04	0.01
S	A	N	P	37.00	0.97	0.04	0.01
S	A	P	N	37.00	0.98	0.02	0.01
S	A	P	P	37.00	0.98	0.04	0.01
S	U	N	N	37.00	0.97	0.04	0.01
S	U	N	P	37.00	0.97	0.04	0.01
S	U	P	N	37.00	0.98	0.03	0.01
S	U	P	P	37.00	0.99	0.02	0.01

Table 37 Summary statistics for reaction time in trials

Summary statistics:							
wordOrder	ambiguity	np1type	np2type	N	Mean	SD	FLSD
O	A	N	N	37.00	484.45	197.41	16.07
O	A	N	P	37.00	435.32	164.68	16.07
O	A	P	N	37.00	482.62	184.68	16.07
O	A	P	P	37.00	433.66	157.27	16.07
O	U	N	N	37.00	465.44	167.91	16.07
O	U	N	P	37.00	437.66	155.49	16.07
O	U	P	N	37.00	442.40	161.16	16.07
O	U	P	P	37.00	424.94	142.27	16.07
S	A	N	N	37.00	436.50	147.15	16.07
S	A	N	P	37.00	427.92	146.34	16.07
S	A	P	N	37.00	442.69	149.04	16.07
S	A	P	P	37.00	427.23	151.57	16.07
S	U	N	N	37.00	438.67	151.13	16.07
S	U	N	P	37.00	433.64	149.12	16.07
S	U	P	N	37.00	437.05	151.31	16.07
S	U	P	P	37.00	420.69	127.54	16.07

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705.
- Arbib, M., Bonaiuto, J., Bornkessel-Schlesewsky, I., Kemmerer, D., MacWhinney, B., Nielsen, F., Oztop, E. (this issue). Action and language mechanisms in the brain: data, models and neuroinformatics. *Neuroinformatics*.
- Baayen, R., Davidson, D., Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- 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.
- Basten, U., Biele, G., Heekeren, H., Fiebach, C. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, 107(50), 21,767–21,772.
- Bates, D., Maechler, M., Bolker, B. (2013). lme4: linear mixed-effects models using S4 classes. <http://CRAN.R-project.org/package=lme4rpackageversion0.999999-2>.
- Bates, E., McNew, S., MacWhinney, B., Devescovi, A., Smith, S. (1982). Functional constraints on sentence processing: a cross-linguistic study. *Cognition*, 11, 245–299.
- Bates, E., Devescovi, A., Wulfeck, B. (2001). Psycholinguistics: a cross-language perspective. *Annual Review of Psychology*, 52, 369–396.
- Bornkessel, I. (2002). *The argument dependency model: a neurocognitive approach to incremental interpretation* (Vol. 28). Leipzig: MPI Series in Cognitive Neuroscience.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821.
- Bornkessel, I., Fiebach, C.J., Friederici, A.D. (2004a). On the cost of syntactic ambiguity in human language comprehension: an individual differences approach. *Cognitive Brain Research*, 21, 11–21.
- Bornkessel, I., McElree, B., Schlesewsky, M., Friederici A.D. (2004b). Multi-dimensional contributions to garden path strength: dissociating phrase structure from case marking. *Journal of Memory and Language*, 51, 495–522.
- Bornkessel, I., Zysset, S., Friederici, A.D., von Cramon, D.Y., Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage*, 26(1), 221–233.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59, 55–73.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Language and Linguistics Compass*, 3(1), 19–58.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013a). Neurotypology: Modelling cross-linguistic similarities and differences in the neurocognition of language comprehension. In M. Sanz, I. Laka, M.K. Tanenhaus (Eds.), *The cognitive and biological basis for linguistic structure: new approaches and enduring themes*, (pp. 241–252). Oxford: Oxford University Press.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013b). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60–76. doi:10.1016/j.bandl.2013.01.010.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (to appear). Competition in argument interpretation: evidence from the neurobiology of language. In B. MacWhinney, A. Malchukov, E. Moravcsik (Eds.), *Competing motivations in grammar and usage*. Oxford: Oxford University Press.
- Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., Schlesewsky, M. (2011). Think

- globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, 117(3), 133–152.
- Bourguignon, N., Drury, J.E., Valois, D., Steinhauer, K. (2012). Decomposing animacy reversals between agents and experiencers: an ERP study. *Brain and Language*, 122(3), 179–189.
- Brouwer, H., Fitz, H., Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143.
- Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer.
- Choudhary, K.K., Schlesewsky, Bickel, B., Bornkessel-Schlesewsky, I. (2010). An actor-preference in a split-ergative language: electrophysiological evidence from Hindi. In *Proceedings from 23rd annual meeting of the cuny conference on human sentence processing*. New York City.
- Coulson, S., King, J.W., Kutas M (1998). ERPs and domain specificity: beating a straw horse. *Language and Cognitive Processes*, 13, 653–672.
- Crocker, M.W. (1994). On the nature of the principle-based sentence processor. In J.C. Clifton, L. Frazier, K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245–266). Hillsdale: Erlbaum.
- Croft, W.A. (2001). *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Dahl, Ö. (2008). Animacy and egophoricity: grammar, ontology and phylogeny. *Lingua*, 118, 141–150.
- Deco, G., Rolls, E.T., Romo, R. (2009). Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, 88(1), 1–16.
- Deco, G., Rolls, E.T., Albantakis, L., Romo, R. (2012). Brain mechanisms for perceptual and reward-related decision-making. *Progress in Neurobiology*.
- 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.
- Demiral, Ş., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: evidence from Turkish. *Cognition*, 106(1), 484–500.
- DeWitt, I., & Rauschecker, J. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, E505–E514.
- Fausey, C., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin and Review*, 18, 150–157.
- Fausey, C.M., Long, B.L., Inamori, A., Boroditsky, L. (2010). Constructing agency: the role of language. *Frontiers in Psychology*, 1(162).
- Federmeier, K.D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505.
- Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1), 1–47.
- Fodor, J.A. (1983). *Modularity of mind. An essay on faculty psychology*. Cambridge: MIT Press.
- Frazier, L. (1987). Syntactic processing: evidence from dutch. *Natural Language and Linguistic Theory*, 5, 519–559. doi:10.1007/BF00138988.
- Frenzel, S., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: a new perspective on the N400–P600 distinction. *Neuropsychologia*, 49(3), 574–579.
- Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, 12(15), 3391–3394.
- Frith, U., & Frith, C.D. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B*, 365, 165–176.
- Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D.Y., Schlesewsky, M. (2006). Linguistic prominence and broca's area: The influence of animacy as a linearization principle. *Neuroimage*, 32, 1395–1402.
- Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, 9, 934–946.
- Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, 20, 18–29.
- Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9), 416–422.
- Hahne, A., & Friederici, A.D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13, 339–356.
- Haupt, F.S., Schlesewsky, M., Roehm, D., Friederici, A.D., Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, 59, 54–96.
- Heekeren, H., Marrett, S., Bandettini, P., Ungerleider, L. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, 431(7010), 859–862.
- Huynh, H., & Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact f-distributions. *Journal of the American Statistical Association*, 65(332), 1582–1589.
- Jonides, J., Lewis, R., Nee, D., Lustig, C. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224.
- Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, 14(2), 129–171.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225.
- Kolk, H.H., Chwilla, D.J., van Herten, M., Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1–36.
- Kretschmar, F. (2010). *The electrophysiological reality of parafoveal processing: on the validity of language-related ERPs in natural reading*. PhD thesis, University of Marburg.
- Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, 1146, 23–49.
- Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463–470.
- Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.
- Lau, E.F., Phillips, C., Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933.
- Leslie, A.M. (1995). A theory of agency. In D. Sperber, D. Premack, A.J. Premack (Eds.), *Causal cognition. A multidisciplinary debate* (pp. 121–141). Oxford: Clarendon Press.
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science: A Multidisciplinary Journal*.

- Lewis, R., Vasishth, S., Dyke, J.V. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Lewis, R.L. (2000). Specifying architectures for language processing: process, control, and memory in parsing and interpretation. In *Mechanisms for language processing*.
- Lotze, N., Tune, S., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia*, 49, 3573–3582.
- MacWhinney, B., & Bates, E. (1989). *The cross-linguistic study of sentence processing*. New York: Cambridge University Press.
- MacWhinney, B., Bates, E., Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 127–50.
- Magnusdotir, S., Fillmore, P., den Ouden, D., Hjaltason, H., Rorden, C., Kjartansson, O., Bonilha, L., Fridriksson, J. (2012). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: a lesion-symptom mapping study. *Human Brain Mapping*.
- Manning, C.D., & Schütze, H. (2000). *Foundations of statistical natural language processing*. Cambridge: MIT Press.
- Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41, 142–151.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–533.
- Martin, A.E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58, 879–906.
- McElree, B. (2006). Accessing recent events. In B.H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego: Academic Press.
- McElree, B., Foraker, S., Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67–91.
- Muralikrishnan, R., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2008). Universal and cross-linguistic influences on the processing of word order and animacy: neurophysiological evidence from Tamil. In *Proceedings from 21st annual CUNY conference on human sentence processing*. Chapel Hill.
- New, J., Cosmides, L., Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104(42), 16,598–16,603.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge: Harvard University Press.
- Nieuwenhuis, S., Aston-Jones, G., Cohen, J.D. (2005). Decision making, the p3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131, 510–532.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Penolazzi, B., Vincenzi, M.D., Angrilli, A., Job, R. (2005). Processing of temporary syntactic ambiguity in Italian who-questions: a study with event-related potentials. *Neuroscience Letters*, 377(2), 91–96.
- Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., Schlesewsky, M. (2008). The role of animacy in the real time comprehension of Mandarin Chinese: evidence from auditory event-related brain potentials. *Brain and Language*, 105(2), 112–133.
- Rauschecker, J. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516–521.
- Rauschecker, J., & Scott, S. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*, 12(6), 718–724.
- Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *NeuroReport*, 15(3), 409–412.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., Schlesewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19, 1259–1274.
- Sassenhagen, J., Kretzschmar, F., Mueller, E., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2013). *Independent components dominating ERP responses to linguistic stimuli also respond to domain-general events*. Manuscript submitted for publication.
- Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: degrees of meaning accessed during sentence comprehension. *Lingua*, 114(9–10), 1213–1234.
- Schlesewsky, M., Fanselow, G., Kliegl, R., Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth, & L. Konieczny (Eds.), *German sentence processing* (pp. 65–93). Dordrecht: Kluwer.
- Schriefers, H., Friederici, A.D., Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34(4), 499–520.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R.M. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). New Jersey: Humanities Press.
- Simon, H.A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467–482.
- Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.) *Handbook of experimental psychology* (pp. 1–49). Wiley.
- Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: an ERP study in Spanish. *Brain and Language*, 120, 108–126.
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Tomlin, R. (1986). *Basic word order: Functional principles*. London: Croom Helm.
- van de Meerendonk, N., Kolk, H.H., Chwilla, D.J., Vissers, C.T.W.M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3, 1211–1224.
- van de Meerendonk, N., Kolk, H.H., Vissers, C.T.W.M., Chwilla, D.J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67–82.
- Vincenzi, M.D. (1991). Filler-gap dependencies in a null-subject language: referential and non-referential whs. *Journal of Psycholinguistic Research*, 20(3), 197–213.
- Vissers, C.T.W.M., Chwilla, D., Kolk, H. (2006). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150–163.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- Vosse, T.G., & Kempen, G.A.M. (2008). Parsing verb-final clauses in German: garden-path and ERP effects modeled by a parallel dynamic parser. In B. Love, K. McRae, V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society (Washington, DC, July 2008)*. Austin: Cognitive Science Society.

- Vosse, T., & Kempen, G. (2009). The unification space implemented as a localist neural net: predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, 3, 331–346.
- Wang, L., Schlesewsky, M., Bickel, B., Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference: evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, 24(7/8), 1180–1226.
- Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36(05), 559–570.
- Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, 47, 360–366.
- Wolff, S., Schlesewsky, M., Hirotani, M., Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, 107, 133–157.
- Zipf, G.K. (1935). *The psycho-biology of language*. Boston: Houghton Mifflin Company.
- Zipf, G.K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.