Combining neurophysiology and formal semantics and pragmatics: the case of the N400

Ralf Naumann[1] and Wiebke Petersen[2]

[1] Heinrich-Heine-Universität Düsseldorf
naumann@phil.hhu.de
[2] Heinrich-Heine-Universität Düsseldorf
petersen@phil.hhu.de

Abstract

We present an outline of how experimental data from neurolinguistics related to one particular ERP-component, the N400, can be analyzed in a probabilistic extension of Incremental Dynamics with frames and situation models. We show that none of the semantic and/or pragmatic properties proposed in the neurolinguistic literature alone can explain the whole range of data. Our own approach is similar to that of Werning et al. in taking the pragmatic dimension seriously by incorporating both the perspective of the speaker and the listener using RSA using a Bayesian model. Probabilities are calculated by using both semantic information which is based on an information ordering on situation models and discourse information which is based on a linking relation between discourse referents.

1 Introduction

An ERP-component is the summation of the post-synaptic potentials of large ensembles (in the order of thousands or millions) of neurons synchronized to an event. When measured from the scalp, continuous ERP-components manifest themselves as voltage fluctuations that can be divided into components. A component is taken to reflect the neural activity underlying a specific computational activity carried out in a given neuroanatomical module. The N400 component is a negative deflection in the ERP signal that starts around 200 - 300 ms post-word onset and peaks around 400ms.

The N400 amplitude on a word w in a context $c = w_1 \dots w_t$ is typically inversely related to its conditional probability given this context: $P(w \mid c)$, [11]. Underlying this relation is a model of online processing according to which at every step during this processing a prediction about the upcoming word is made guided by the probability distribution $P(w \mid c)$ (see [11] for an overview). The N400 amplitude shows a gradient effect. It is smallest for the most predicted words, intermediate for the words with moderate predictability and larger for words with the lowest predictability (see e.g. [18] and [13]). This conditional probability can be measured either by the human judged cloze probability or by the information-theoretic notion of surprisal. Given an initial sequence of words $w_1 \dots w_{t-1}$, w_t can be viewed as a random variable. Its surprisal (or self-information) is defined as follows:[1] $surprisal(w_t) := -logP(w_t \mid w_1 \dots w_{t-1})$. If defined in this way surprisal of a word is typically directly related to the word's N400 amplitude, [9].

The context $w_1 \dots w_{t-1}$ must not be restricted to a single clause or sentence because the wider context can have an influence on the modulation of the N400 amplitude. For example, the probabilities for the target words in the sentences "The peanut was salted/in love" are inverted, if the sentence is not presented in isolation but in the context of a comic or a fiction story in which the peanut is ascribed typical human properties like being able to sing and dance. As an

[1] The base of the logarithm is an arbitrary scaling factor.

effect, a property like 'being salted' is now highly unlikely if not even impossible. By contrast, the property of being in love now receives a high probability. This inversion of probabilities is reflected in the N400 amplitude: for 'salted' this amplitude was enhanced compared to that for 'in love', [16].

The two most prominent interpretations of the underlying neuro-cognitive function of the N400 are the integration and the retrieval view. On the integration account, the N400 amplitude 'indexes the effort involved in integrating the word meaning of the eliciting word form with the preceding context, to produce an updated utterance interpretation', [7]. On the retrieval/access account 'the N400 amplitude reflects the effort involved in retrieving from long-term memory conceptual knowledge associated with the eliciting word which is influenced by the extent to which this knowledge is cued (or primed) by the preceding context, [7]. What is left open by the above characterization is which properties of words and the context underly the N400 amplitude. Three prominent properties that have been suggested are (i) semantic features, (ii) plausibility, and (iii) semantic similarity.

Evidence for semantic features as being correlated with the N400 amplitude comes from the fact that the correlation between cloze probability and the N400 amplitude is not monotone.

(1) They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of palms/pines/tulips. [8]

In (1) 'pine' but not 'tulips' comes from the same semantic category 'tree' as the best completion 'palms'. Though 'pines' and 'tulips' have the same low cloze probability (< 0.05), their N400 amplitudes differ. Within category violations (pines) elicit smaller N400 amplitudes than between category violations (tulips). Federmeier & Kutas argue that this result suggests that it is feature overlap like being tall or having a similar form that affords within category violations a processing benefit relative to between category violations, [8, p.485].

However, feature overlap with the best completion is not without exceptions, as shown by the following examples.

(2) a. A huge blizzard swept through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big jacket in the front yard, [14].
b. The wreckage of the sunken ship was salvaged by the victims ... [17].

Though the critical words share few semantic features with the best completions (snowman, divers), either no, (2-b), or only a small N400 effect, (2-a), is observed.

A second candidate is plausibility which can be quantized by offline rating tasks using, e.g., a Likert scale. Plausibility is often related to the integration view of the N400. The less plausible a resulting interpretation is the more difficult must it have been to integrate the critical word in the preceding context. Evidence for the role of plausibility comes from the fact that in the Federmeier & Kutas study best completions elicited the smallest N400 amplitude and the highest plausibility ratings. Between category violations elicited the largest N400 amplitudes and got the lowest plausibility ratings. Within category violations were intermediate on both variables, [8, p.486]. However, this monotone relation no longer holds if contextual constraint is taken into account. Most importantly, in low-constraint contexts the plausibility for within category violations is significantly higher compared to high-constraint contexts. By contrast, the N400 amplitudes are significantly different in the opposite direction. The more plausible within category violation in low-constraint contexts have a higher N400 amplitude than the less plausible within category violations in a high-constraint context. Furthermore, in semantic illusion data as given in (3) no N400 effect is observed although the sentence has an implausible

interpretation.

(3) The fox that on the poacher hunted ....

A third candidate is semantic similarity. On this account the N400 amplitude is modulated by the degree to which a critical word in a target sentence is semantically related to the words preceding it in the context. For example, in the peanut example one has the 'being in love' has a higher semantic similarity to words like 'dancing' and 'singing' than to 'being salted'. One way of quantifying semantic similarity is to use Latent Semantic Analysis. On this account pairwise term-to-document semantic similarity values (SSVs) are extracted from corpora (see [13] for an application). Semantic similarity underlies the Retrieval-Integration model of [20]. One of its strengths is that it can explain semantic illusion data as given in (3). As there is a semantic relation between the arguments preceding the verb ('fox', 'poacher') and the verb itself ('hunted') no N400 effect is expected for the verb.

However, similar to both the notions of semantic feature overlap and plausibility, there are counterexamples to the thesis that the N400 amplitude is (monotonically) related to the corresponding LSA value. Kuperberg et al., [12] showed that the degree of causal relationship in three-sentence scenarios with matched SSVs influences the N400 amplitude: highly related < intermediately related < causally unrelated. The authors conclude that it is the situation model constructed from the context (message-level meaning) that influences semantic processing of the critical word and not semantic relatedness. Similarly, [13] could show an influence of high- versus low-constraint contexts on the N400 amplitude for controlled SSVs.

Let us summarize the findings of this section: The modulation of the N400 amplitude is sensitive to (i) semantic feature overlap and not simply to words (for a similar argument, see [13]). Exceptions are cases in which the sort of the word actually found does not share semantic features with the best completion, but with the preceding context. (ii) Plausibility judgements accounts for the preceding context, but fail to explain semantic illusion data. Finally, (iii) semantic similarity as defined by LSA abstracts away from thematic roles and thus captures semantic illusion data. However, it is not restrictive enough, as it does not account for causal relatedness and degree of context constraint.

An alternative is proposed by Werning and colleagues, [21], [22]. They start from the assumption, already discussed in section 1, that at any moment in a communicative situation a comprehender generates a probabilistic prediction about how a sentence or a discourse uttered by a speaker will most likely be continued, [21, p. 3504]. This communicative act is goal-directed, i.e. the speaker wants to describe a particular situation. Hence, he will choose a context c which makes the referent denoted by w (highly) relevant. They define $P(w \mid c)$ not in terms of a single property. They rather follow the rational speech act model (RSA) and assume a Bayesian model, see e.g. [6]. Such a model allows comprehenders to update their priors regarding a word w following a context c with pragmatic considerations on speakers' intentions thereby arriving at a probabilistic prediction of w. The conditional probability $P(w \mid c)$ is defined as the product of a probabilistic semantic factor, given by the prior, and a pragmatic (discourse) component that is represented by the likelihood term.

(4) $P(w \mid c) \propto P(w) \cdot P(c \mid w)$.

The prior is a function of the semantic similarity between w and c and/or another word w' in c and is defined using LSA. This reflects overall statistical co-occurrence patterns and hence statistical regularities. The likelihood term models the pragmatic dimension. Given that a speaker wants to convey information about the referent of a word w he will choose a context c that makes the occurrence of this referent relevant (plausible). Hence $P(c \mid w)$ strictly increases

with the relevance (plausibility) of c for w. In their empirical studies they showed that relating the modulation of the N400 amplitude to (4) yields empirical better results than relating this modulation to either only semantic similarity (prior) or relevance (plausibility) (likelihood). Problematic for this approach are semantic illusion data and examples like (2) for which there is no or only an attenuated N400 effect despite the fact that there is no semantic feature overlap between the critical word and the best completion. In the case of semantic illusion data the semantic similarity between the verb and the two preceding arguments is the same in the expected and the switched thematic role variants because switching thematic roles does not affect the SSV value. However, they differ w.r.t. relevance (plausibility) due to the difference in thematic role assignments. As a result, an N400 effect for the switched roles variants compared to the expected role assignments should be observed. Similarly, for the data in (2) the semantic similarity between the critical word and the expected best completion is nearly identical (e.g. in the case of (2-b) one has SSV(divers, context)=0.22 and SSV(victims, context) = 0.18). By contrast, the difference in plausibility is significant: 6.3 vs 2.9 on a 7-point Likert scale. Hence, an N400 effect for 'victims' relative to 'divers' is expected, contrary to what was observed. Given the results of the first section, these problems reflect the failure of plausibility being an underlying factor of the modulation of the N400 amplitude. Relevance/plausibility applies at the propositional/discourse level and hence at the level of event structures. This raises the question whether there is another property, possibly related to a different semantic object, that underlies this modulation.

2 Towards an RSA account on language processing that respects neurophysiological findings

If quantized by LSA, semantic similarity is based on semantic relationships between words and concepts, including (inferential) schema-based relationships. It is insensitive to word order and both syntax and thematic relation. An example of such inferential relationships is the qualia structure in the Generative Lexicon. It links a sort of objects, say cream, to a particular action (or a set of actions) that specifies the function or purpose of objects of this sort. For a comprehender who processes the verb denoting this action the interpretative task is to relate the corresponding discourse referent to the discourse referent of the noun to whose qualia structure the action belongs. This suggests that the N400 is sensitive to establishing such relations between discourse referents. This hypothesis raises two questions: (i) is there direct evidence that the N400 is sensitive to such relations between objects?, and (ii) How can these relations be defined? Consider the examples in (5).

(5) a. Peter hatte einen langen Tag und wollte ein Bier. Die Kneipe war bis Mitternacht geöffnet/Das Essen war bereits auf dem Tisch, [7].
Peter had a long day and wanted a beer. The bar was open till midnight./ The meal was already on the table.

b. Tobias besuchte einen Dirigenten/ein Konzert in Berlin/unterhielt sich mit Nina. Er erzählte, daß der Dirigent sehr beeindruckend war, [3].
Tobias visited a concert/a conductor in Berlin/talked to Nina. He said that the conductor was very impressive.

In contrast to factive verbs, the existence of the direct object of non-factive verbs like 'want' is neither presupposed nor asserted. 'Want' raises a particular question under discussion: 'Did the actor got what he wanted?' This question triggers the expectation that this question will be

answered. A bar is a paradigmatic place where one gets beer (or not if, e.g. it is already closed) and where one can go if one wants a beer. By contrast, a meal can be served without any beverages and, in addition, a beer being served is only one among many possibilities. Delogu et al. found an N400 effect for 'Essen' compared to 'Kneipe'. The examples in (5-b) involve bridging inferences. Burkhardt found an attenuated N400 effect for bridged DPs (Konzert - Dirgient) and an enhanced effect for new DPs (Nina - Dirigent) compared to the given DP (Dirigent - Dirigent). She calls the general phenomenon 'discourse linking'.

Examples like those in (2) show that discourse linking cannot be restricted to (single) event structures. Rather, it is related to situation models. Such models go beyond the propositional content conveyed by the words in a context and which essentially involve world knowledge, [23] see [1] on learning such models from texts]). For example, in (2-a) a wintery scene and in (2-b) a ship wreckage scenario is described. Situation models basically are complex events. Since situation models comprise sequences of events, predictions are not restricted to single events. For example, given a context that specifies a situation model whose prototypical realization consists of the action sequence $e_1 \dots e_r$ and in which the initial sequence $e_1 \dots e_k$ has been introduced, predictions are possibly related to any of the events $e_{k+1} \dots e_r$ and objects participating in them. In the wintery scenery in (2-a) the jackets are expected because they are related to the children in a (yet to be introduced) state of wearing which is a background state that constantly holds while the children were playing outside. As a second example, consider (6), which is an example of semantic illusion data.

(6) The restaurant owner forgot which waitress the customer had ...

Given a restaurant scenario, in which a waitress and a customer have already been introduced, actions like 'serve', 'ask', 'order' or 'pay' are expected. In this case these actions are not only related to the restaurant scenario but also to the current event whose sort is still unknown. Hence, situation models possibly set up predictions to objects in event structures that have not yet been introduced.

We hypothesize that the modulation of the N400 amplitude is sensitive to (i) semantic similarity between situation models and (ii) the way situation models are related by discourse linking. Let us relate this to the RSA approach (see e.g. [6]). The main insight underlying the RSA model is that a listener not only uses the literal meaning possibly enriched by world and script knowledge but also takes into account that a speaker chooses an expression in such a way that he is able to infer the intended referent of the expression. Hence, speaker and listener recursively reason about each others' goals to arrive at pragmatically enriched meanings. Formally, one has that a speaker S_1 chooses a term t to (soft-max) optimize expected utility given a meaning (referent) r: $S_1(t \mid r) \propto e^{\lambda(U_1(t \mid r))}$ (λ is the gain on the speaker's softmax decision rule). A literal listener interprets utterances literally without reasoning about the speaker. He has a prior distribution over referents and uses Bayesian inference to (eliminatively) update her belief about the intended referent given the utterance's literal meaning. A pragmatic listener L_1 then reasons about S_1 by inverting S_1's model using Bayes' rule in order to infer the referent r given utterance u, where $P(r)$ is the prior probability over referents: $L_1(r \mid t) \propto P(r)S_1(t \mid r)$. In our application t and r are situation models sm and sm'. The task, therefore, is to define $P(sm)$ and the utility function $U_1(sm, sm')$. In order to solve this task, one has to define the meaning of common nouns and verbs in terms of semantic features. This will be done by a decompositional analysis. Second, probabilities have to be defined in terms of such decompositional structures. To this end, frames and their properties will be introduced.

In order to account for the modulation of the N400 amplitude at the semantic feature level, common nouns and verbs are not interpreted as sets of objects, either individuals or events.

Rather, they are interpreted as sets of pairs $\langle o, f\rangle$ consisting of an object and a frame. Frames are elements of a separate domain D_f of frames. Each frame is related to a particular object (an individual or a (complex) event) as its root and is a partial description of that object in a particular world. Being a partial description of an object, a frame is linked to a relational structure that is built by (finite) chains of attributes. This link is captured by a function θ which maps a frame to a set of relations. For a given object, its associated frame stores information got during a discourse so far as well as world knowledge. Besides the domain D_f, there are the domains D_i of individuals, the domain D_e of events and the domain D_w of possible worlds.

At the discourse level we use Incremental Dynamics [19] enriched with frames [15]. We extend our approach in [15] by set-valued frames for the current situation. Situation models or possibilities are triples $sm = \langle c_{sm}, f_{sm}, w_{sm}\rangle$ consisting of a stack c_{sm}, a current situation frame f_{sm} and a world w_{sm}. A particular stack position is a pair $\langle o, f\rangle$ with $o \in D_i \cup D_e$ and $f \in D_f$. An information state is a set of possibilities. Every situation frame f_{sm} has an attribute ACTIONS whose value is the set of actions (events) occurring in this scenario (denoted by $a(f_{sm})$). A second attribute is PARTICIPANTS whose value is a set of individuals $p(f_{sm})$. Each element of this set is related to at least one action or one other participant, the set of these pairs $pr(f_{sm})$ is the value of the attribute PARTICIPANCY_RELATION. The value of the attribute ORDER is a set $o(f_{sm})$ of pairs of events that preorders the value of ACTIONS attribute. Situation frames are sorted by SM which are sorts of complex events like 'wintery scenario' or 'restaurant scheme'.

In order to capture the sensitivity of the modulation of the N400 amplitude to both semantic features and world knowledge probabilities need to be defined not at the level of words but at the frame level and particularly at the level of situation frames. Whenever a new word is processed, the current situation model is updated. In the case of a newly introduced discourse referent, the set values in the situation frame are extended and the stack is prolonged by one position (see [15] for details). In case of an anaphora there exists already a corresponding object frame pair in the stack of which the frame is updated, i.e. refined, by the newly gained information (see again [15] for details). On the level of ordinary frames the refinement can be defined in terms of the following information ordering $\sqsubseteq$ on frames. Let $\theta(f) = \{R_1, \ldots R_n\}$ be the set of chains associated with f. One has $f \sqsubseteq f'$ if (i) f and f' have the same root and (iii) $\theta(f) \subseteq \theta(f')$. When taken together, this means that f' possibly contains information about more attributes or f' possibly contains more specific information about the values of attributes. For set-valued situation frames and full situation models the information order is as follows:

(7) a. $f_{sm} \sqsubseteq f_{sm'}$ iff $a(f_{sm}) \subseteq a(f_{sm'})$ and $p(f_{sm}) \subseteq p(f_{sm'})$ and $pr(f_{sm}) \subseteq pr(f_{sm'})$ and $o(f_{sm}) \subseteq o(f_{sm'})$.

b. $sm \sqsubseteq sm'$ iff (i) $w_{sm} = w_{sm'}$, (ii) $f_{sm} \sqsubseteq f_{sm'}$ and (iii) $\forall o$: if $\langle o, f\rangle \in c_{sm}$ and $\langle o, f'\rangle \in c_{sm'}$ then $f \sqsubseteq f'$.

Thus a situation model sm' extends or refines a situation model sm if both belong to the same world and the attribute values of $f_{sm'}$ are supersets of the corresponding values of f_{sm} and for all objects belonging to sm there corresponding frames stored in the stack c_{sm} are possibly refined in sm'. An update is a move along the information hierarchy: Let the context be given by the words $w_1 \ldots w_{t-1}$ with corresponding frames $f_1 \ldots f_{t-1}$ resulting in a situation model $sm_{w_1^{t-1}}$. The meaning of the word w_t is a context change potential. In the present context this is the change it brings about with respect to $sm_{w_1^{t-1}}$ yielding the updated situation model $sm_{w_1^t}$. The contribution of w_t is an object frame pair $\langle o, f_t\rangle$. The operation of updating sm_1^{t-1} with $\langle o, f_t\rangle$ will be denoted by $\oplus$: $sm_1^t = sm_1^{t-1} \oplus \langle o, f_t\rangle$. Let's assume that w_t introduces a new

discourse referent $o \notin p(_{sm})$. Then it extends the stack, $c_{sm'} = c_{sm} \frown \langle o, f_t \rangle$, and introduces a new participant, $p(f_{sm'}) = p(f_{sm}) \cup \{o\}$, that is linked to at least one other participant or action, $\exists o' \in p(f_{sm'}) \cup a(f_{sm'}) : pr(f_{sm}) \cup \{\langle o, o' \rangle\} \subseteq pr(f_{sm'})$. The case for a new action discourse referent $o \notin a(f_{sm})$ is similar. If w_t denotes an object o that is already on the stack, $\langle o, f \rangle c_{sm}$ then the corresponding frame is updated by the information f_t via unification: $\langle o, f \sqcup f_t \rangle \in c_{sm'}$.

Next we define the prior for a situation model sm given a situation model sm'. In order to do so, we need to abstract away from the concrete worlds and the concrete participants. Two situation models sm and sm' are alphabetic variants of each other, $sm \approx sm'$, iff there are bijections $\phi : p(sm) \to p(sm')$ and $\phi : a(sm) \to a(sm')$ such that $f_{sm'} = \phi(f_{sm})$ and $c_{sm'} = \phi(c_{sm})$. Thus alphabetic variants differ only in their worlds, their participants and their actions, but not in their structure and thus not in their general properties. Given a situation model sm resulting from the processing of a context $w_1 \dots w_{t-1}$, the prior for a possible extension of sm by the frame f_t is

$$P(sm \oplus f_t) = \frac{|\bigcup_{sm' \approx sm \oplus f_t} \uparrow sm'|}{|\bigcup_{f \in D_f \wedge sm' \approx sm \oplus f} \uparrow sm'|}$$

with $\uparrow sm$ being the filter of sm, i.e. $\uparrow sm = \{sm' | sm \sqsubseteq sm'\}$.

$P(sm \oplus f_t)$ will be used as the prior. Note that the prior is calculated without taking thematic role information into account. The reason why predictions are based on semantic features and not on information based on thematic roles has to do with ambiguity triggered by literal meaning which considerably increases processing load with the effect that it is not possible to arrive at a stable representation without making use of predictions, [5]. Semantic processing in the brain is done in a left-to-right fashion. This makes it necessary to rethink the way arguments are related to verbs (see [2] for evidence and discussion). Following [4] and [2], we assume the following structure of a DP: $[[DetN]_{DP_1}[TR]]_{DP_2}$ in which the contribution of sortal and thematic role information are separated. On this interpretation the assignment of a thematic role can be taken as a non-deterministic operation which introduces branching (at least for case-less languages like English or Dutch). The crucial difference between the two kinds of information is that for a listener only the sortal information is directly given whereas thematic role information has to be inferred.

Recall that on a pragmatic-discourse oriented perspective texts are goal-oriented. In our approach the global aim of the speaker is to describe a particular situation. To this end, he will locally introduce objects that are part of this situation and attribute properties to them. He will therefore choose a (prior) context that makes the mention of these objects (most) likely. Such a mention is highly probable if the context already contains this object as an element. Next come objects that are (directly) related by an attribute to objects already introduced in the context since they, at least indirectly, extend information about objects already introduced. This has the effect of making the text coherent by maximizing anaphoricity. We hypothesize that these probabilities are related to particular accessibility relations between situation models that are defined in (8).

(8) a. $sm \underrightarrow{\text{Id}} sm'$ iff (i) $f_{sm} \sqsubseteq f_{sm'}$ and (ii) $a(sm) \cup p(sm) = a(sm') \cup p(sm')$.
b. $sm \underrightarrow{\text{BI}} sm'$ iff (i) $f_{sm} \sqsubseteq f_{sm'}$ and (ii) $\forall o' \in (a(sm') \cup p(sm')) \setminus (a(sm) \cup p(sm)) : \exists \langle o, f \rangle \in c_{sm} \exists f' \exists R$ with $f \sqsubseteq f'$ and $R(f')(o)(o')$.
c. $\underrightarrow{\text{Id}} \cup \underrightarrow{\text{BI}} \subseteq \underrightarrow{\text{DL}}$; $\underrightarrow{\text{DL}} \subseteq \sqsubseteq$.

$\underrightarrow{\text{Id}}$ requires the two situation models to have the same participants and events. Hence, this relation captures the case of given DPs. $\underrightarrow{\text{BI}}$ requires that for each object that belongs to sm'

but not to *sm* there is an object belonging to *sm* and a relation in a minimal extension of the frame of this object that links the two objects. Thus, this relation captures the case of bridged DPs. Finally, discourse linking, $\underrightarrow{\text{DL}}$, comprises these two relations and is itself a subrelation of $\sqsubseteq$. This can be tested by using a procedure based on the next-mention bias used by [10] for anaphoric relations involving pronouns. This bias will be largest for objects already introduced, in particular if they are related to the topic. Similarly, bridged DPs answering a QuD as in the example (5-a), will get a high probability.

How is discourse linking based on the accessibility relations in (8) related to the information conveyed by a word? Recall that the aim of a speaker is to describe a particular situation. At each point in processing there is some degree of uncertainty for the listener about which situation model is described. This uncertainty decreases by getting to know the frame of the next word and the extent of this decrease corresponds to the information contained in that frame. According to discourse linking, this decrease should be related to making the discourse coherent by maximizing anaphoricity. This will be the case if the interpretation of a DP does decrease uncertainty only minimally. Let us make this formally precise. Recall that in RSA the speaker's utility function must be related to the information conveyed by a word in an utterance. This information is often quantized by surprisal discussed in the first section. However, as also shown there, this metric can neither account for semantic illusion data nor for the examples in (2) where there is no feature overlap between the critical word and the expected best completion. As alternative to surprisal, we use entropy reduction which is based on the notion of (n-step) entropy and which is directly related to uncertainty about a referent. For $S_{f_{comp}}$ the set of all completions, the listener wants to know which situation model *sm* is described by determining a frame $f_{sm} \in S_{f_{comp}}$. The uncertainty about f_{sm} is defined as the entropy of the probability distribution over $S_{f_{comp}}$: $H(S_{f_{comp}}) = -\sum_{f_{sm} \in S_{f_{comp}}} P(f_{sm}) log(P(f_{sm}))$. When the first t words of a sentence have been processed, the probability distribution over $S_{f_{comp}}$ has changed from $P(S_{f_{comp}})$ to $P(S_{f_{comp}} \mid f_{w_1^t})$ with $f_{w_1^t}$ being the frame got after the first t words. The corresponding entropy equals $H(S_{f_{comp}}, f_{w_1^t}) = -\sum_{f_{sm} \in S_{f_{comp}}} P(f_{sm} \mid f_{w_1^t}) log P(f_{sm} \mid f_{w_1^t})$. The amount of information that the next word w_{t+1} gives about the random variable $S_{f_{comp}}$ is defined as the reduction in entropy due to that word: $\Delta H(S_{f_{comp}}; f_{w_{t+1}}) = H(S_{f_{comp}}; f_{w_1^t}) - H(S_{f_{comp}}; f_{w_1^{t+1}})$. Let's apply this to some examples from discourse linking. Consider first the case of a given DP like 'the conductor' in (5-b). This kind of DP does not exclude any extensions that were possible before this DP was encountered because the information related to this DP was already known in the input information state. For bridged DPs, this will in general not be the case because some extensions are excluded by establishing a linking relation that was not known before. Take, for example, the case of the jackets in (2-a). This excludes situations in which the children were wearing coats or ski suits. A bridged DP need not always exclude extensions. In the case of the ship wreckage, e.g., it is not excluded that there were other material casualties besides the victims. This example shows that entropy reduction is more fine-grained than the subrelations of $\underrightarrow{\text{DL}}$ because it allows for distinctions in the $\underrightarrow{\text{BI}}$ relation. By contrast, a new DP leads to an increase in entropy because in general it cannot be directly linked to an object already introduced. For example, in the case of the conductor about whom Tobias talks to Nina there is no direct link relating the former to the latter two objects. What is missing, e.g., is a discourse referent that establishes such a relation. This can be a concert that Tobias attended. When taken together, one gets that a speaker chooses a context in such a way that the reduction in entropy is minimal for the object he wants to talk about next. Entropy reduction is calculated relative to the ordering $\underrightarrow{\text{DL}}$. Hence, the contribution of f_t in determining an element of $S_{f_{comp}}$ is always related to the transition from *sm* to *sm*$'$ triggered by f_t relative to

this ordering. This contribution is denoted by $\oplus_{\overrightarrow{\text{DL}}}(sm, sm', f_{w_{t+1}}))$ When taken together, the utility function U_1 is defined as follows: $U_1(sm \mid sm') := -\Delta H(S_{f_{comp}}; \oplus_{\overrightarrow{\text{DL}}}(sm, sm', f_{w_{t+1}}))$.

Let us next discuss some examples used above. We will begin with the two examples in (2). Both 'jackets' and 'victims' can be linked to an object that is an element of the current situation model. The jackets are the value of the CLOTHES attribute and the victims are the value of the CASUALTIES attribute. As already stated above, they differ w.r.t. the information they provide. Whereas 'victims' does not exclude other casualties, 'jackets' excludes other sorts of outer clothing like coats for example. Similarly, the prior for 'victims' will in general be greater than that for 'jackets' due to the fact that the children can wear other clothes. Comparing 'victims' with the most expected 'divers', one gets: in both cases a new object is introduced into the situation model. Similar to 'victims', 'divers' does not exclude any sort because salvaging a shipwreck requires objects of this sort. The priors will also be equal so that no N400 effect is expected. For the example (6) one has: The prior for an event of serving is the same for a restaurant scenario in which a waitress and a customer have been introduced. Note that this only holds if thematic role assignments are not taken into account. Similarly, the action of serving can equally be linked to objects of sort 'waitress' and 'customer' as actions in which objects of this sort are involved. Since the action is identical for both assignments of thematic roles, entropy reduction does not differ. Consider next example (1). The critical words differ with respect to their prior probabilities. Given the preceding context, 'palm' are more expected because they satisfy more of the (soft) constraints imposed on the object planted, e.g. that its typical origin are the tropics. For linking at the discourse level, one gets: in the input information state an event of planting has already been introduced which expects objects of sort 'plant'. This constraint is satisfied by all three sorts which occur in the position of the critical word. They do not differ w.r.t. entropy reduction because they all paradigmatically exclude the other possibilities.

3 Acknowledgments

We would like to thank Harm Brouwer for illuminating discussions on the neurolinguistic data and the relation to the Retrieval-Integration model. Markus Werning provided helpful discussions of his own approach as well as of possible ways to combine neurolinguistic data with formal semantic theories. The research was supported by the German Science Foundation (DFG) funding the Collaborative Research Center 991.

References

[1] Simon Ahrendt and Vera Demberg. Improving event prediction by representing script participants. In *Proceedings NAACL 2016*, pages 546–551. ACL, 2016.

[2] Oliver Bott and Wolfgang Sternefeld. An event semantics with continuations for incremental interpretation. *Journal of Semantics*, 34(2):201–236, 2017.

[3] Petra Burkhardt. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159 – 168, 2006.

[4] Lucas Champollion. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy*, 38(1):31–66, Feb 2015.

[5] Herbert Clark. *Using language*. Cambridge University Press, 1996.

[6] Judith Degen, Michael Henry Tessler, and Noah D. Goodman. Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Proc. of CogSci 2015*, page 548–553.

[7] Francesca Delogu, Harm Brouwer, and Matthew W. Crocker. Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, 135:103569, 2019.

[8] Kara D. Federmeier and Marta Kutas. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41:469, 1999.

[9] Stefan L. Frank, Leun J. Otten, Giulia Galli, and Gabriella Vigliocco. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1 – 11, 2015.

[10] Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey Elman. Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes*, 53:1–20, 2016.

[11] Gina Kuperberg and T. Florian Jaeger. What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, 31(1):32–59, 2016.

[12] Gina Kuperberg, Martin Paczynski, and Tali Ditman. Establishing causal coherence across sentences: an ERP study. *Journal of Cognitive Neuroscience*, 23(5):1230–1246, 2011.

[13] Gina R. Kuperberg, Trevor Brothers, and Edward W. Wlotko. A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 0(0):1–24, 0.

[14] Ross Metusalem, Marta Kutas, Thomas P. Urbach, Mary Hare, Ken McRae, and Jeffrey L. Elman. Generalized event knowledge activation during online sentence comprehension. *Journal of memory and language*, 66 4:545–567, 2012.

[15] Ralf Naumann and Wiebke Petersen. Bridging inferences in a dynamic frame theory. In Alexandra Silva, Sam Staton, Peter Sutton, and Carla Umbach, editors, *Language, Logic, and Computation*, pages 228–252, Berlin, Heidelberg, 2019. Springer Berlin Heidelberg.

[16] Mante S. Nieuwland and Jos J. A. Van Berkum. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7):1098–1111, 2006.

[17] Martin Paczynski and Gina Kuperberg. Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4):426–448, 2012.

[18] Dianne E. Thornhill and Cyma Van Petten. Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3):382 – 392, 2012.

[19] Jan van Eijck. On the proper treatment of context in nl. In Paola Monachesi, editor, *Computational Linguistics in the Netherlands 1999. Selected Papers form the Tenth CLIN Meeting*. UILOTS Utrecht, 1999.

[20] Noortje J. Venhuizen, Matthew W. Crocker, and Harm Brouwer. Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, 0(0):1–27, 2018.

[21] Markus Werning and Erica Cosentino. The interaction of bayesian pragmatics and lexical semantics in linguistic interpretation: Using event-related potentials to investigate hearers' probabilistic predictions. In *Proceedings of CogSci 39*, pages 3504–09, 2017.

[22] Markus Werning, Matthias Unterhuber, and Gregor Wiedemann. Bayesian pragmatics provides the best quantitative model of context effects on word meaning in EEG and cloze data. In *Proceedings of CogSci 41*, pages 3085–91, 2019.

[23] Rolf Zwaan and Gabriel A. Radvansky. Situation models in language comprehension and memory. *Psychological Bulletin*, 123:162 – 185, 1998.