Semantic predictions in natural language processing, default reasoning and belief revision

Ralf Naumann and Wiebke Petersen

Institut für Sprache und Information
Universität Düsseldorf
Germany

Abstract. Formal semantic theories are designed to explain how it is possible to produce and understand an infinite number of sentences on the basis of a finite lexicon and a finite number of composition rules. According to this architecture, language comprehension completely proceeds in a bottom-up fashion only driven by linear linguistic input thereby leaving no room for a predictive component which allows to make expectations about upcoming words. This is in stark contrast to neurophysiological research in the past decades on online semantic processing which has provided ample evidence that prediction plays indeed an indispensable role in language comprehension (the brain as a *prediction machine*, [Ber10]). In this article, we present an extension of formal semantic theory that allows to make predictions of upcoming words. The basic intuition is: predictions are based on incomplete information. Drawing (defeasible) conclusions based on such information can be modeled by default reasoning. Since predictions can go wrong, a second strategy for retracting wrong guesses is needed in order to to integrate (unexpected) words into the prior context. This is modeled by belief revision. We model both processing stages, making predictions about upcoming words and integrating them into the prior context, and relate the models to the empirical findings in neurophysiological research.[1]

Keywords: default logic, modal logic, cognitive semantics, system Z, N400, late positivity

1 The brain as a prediction machine

In formal semantic theories meaning is taken to be a relation between language and the external world (or reality). This relation is defined inside a logical theory, e.g. some form of type logic, using notions like ‘reference’, ‘satisfaction’ and ‘truth’. On this view the main goal of natural language semantics is a definition

[1] We thank our anonymous reviewers who helped to improve this paper by providing an elaborated and constructive feedback to former versions. The research was supported by the German Science Foundation (DFG) funding the Collaborative Research Center 991.

of the truth for sentences in a natural language. This goal is achieved by giving a recursive and compositional analysis of the well-formed expressions of a language. Based on a finite lexicon and a finite set of composition rules, it then becomes possible to both produce and parse an infinite set of sentences none of which needs to be stored in the brain. This characterization is still valid for dynamic approaches like DRT or DPL in which the notion of truth is replaced by that of a relation between (information) states.

From a psycholinguistic or neurophysiological point of view the concept of meaning endorsed in formal semantics is quite unsatisfactory since it completely leaves out the question of how language is processed in the brain. A prime example that has emerged during the last three decades both in behavioral and electro-physical research are predictions or expectations of upcoming words in a given context.[2] Consider the example in (1) taken from [FK99, 469].

(1) Getting himself and his car to work on the neighboring island was time consuming. Every morning he drove for a few minutes and then boarded the . . . .

When asked, most people end the second sentence with the word ‘ferry’. This behavior is remarkably robust across individuals and it is empirically defined in terms of a word’s cloze probability[3] in a given (sentential) context. For example, in (1), ‘ferry’ has highest cloze probability (CP) and is therefore the *best completion* (BestComp). Since none of the individual words in (1) is strongly semantically related to ‘ferry’, it seems most likely that the context preceding ‘. . . ’ together with world knowledge is used during language processing to pre-activate semantic properties which best apply to (the concept expressed by) ‘ferry’ but not to the same degree to other vehicles like gondolas or airplanes. On this interpretation, both world knowledge and context play a crucial role in setting up semantic properties on the basis of which an expectation (or prediction) for an upcoming word is formed.

According to Baggio and Hagoort, examples like (1) show that formal semantics ‘is by design insensitive to differences between words of the same syntactic category denoting objects of the same type’, [BH11, 1343]. Their own example is (2).

(2) Last Friday the cruiser Arberia entered the *port/hippodrome* of Trieste.

They argue that the difference between the two continuations after ‘entered the’ must be semantic in nature because pragmatic deviance like the violation of a Gricean conversational maxim does not occur (if one assumes that both sentences

[2] ‘Prediction’ must not be understood as a conscious or strategic process. Rather, prediction is understood as the unconscious activation of semantic properties of upcoming words prior to their occurrence, [FK99, 487].

[3] Cloze probability: participants in an offline norming task are presented sentence frames like that in (1) and are asked to fill in the dots with the first word that comes to their mind. The proportion, ranging from 0 to 1, of respondents supplying a particular word is defined as the cloze probability of this word in that context.

are false at speech time). In addition, the difference has nothing to do with the way the world looks like.

However, note that Baggio and Hagoort's argument is based on the implicit assumption that the problem arises only at the level of integration/composition. After 'port' or 'hippodrome' have been semantically recognized, they have to be integrated or combined with (the semantic representation of) the previous context. For 'port', being a best completion, this should pose no problems whereas for 'hippodrome' this integration should be much more difficult, if not impossible, given that this word is not only semantically unrelated but almost semantic anomalous to the semantic properties of the context. Though integration and prediction are closely related, the problem of how predictions and/or expectations can be represented in formal semantic theories cannot be reduced to simply incorporating it into the integration/composition mechanism.

If prediction (or expectation) is understood in the sense that it is based on the pre-activation of semantic features of words which are not yet presented to the comprehension system, the problem of combining or of integrating that word with the current semantic representation does arise only at a second stage. In a first stage, the semantic features of the expected upcoming, not yet presented, word are activated simultaneously (or in parallel) with the semantic features of words that have already been recognized and combined with the prior context.[4] Thus, there must be a separate mechanism which makes it possible to deduce semantic features (σ) from information that is already part of the semantic representation (τ) of the prior context and world knowledge (τ') stored in Long Term Memory (LTM). Then, using τ and τ', σ is deduced. Prediction is closely related to integration. Since predictions are risky – they can go wrong – there needs to be an additional (or subsequent) mechanism that deals with wrong guesses by explaining how they can be retracted. Exactly at this point prediction becomes related to integration/composition. Predicted semantic features are used to build up a semantic representation of the upcoming word, which eventually is integrated with the prior context. If a prediction turns out to be wrong because a non-expected word is encountered, integration is successful only if the wrong guesses are first retracted because otherwise combining the predicted with the actual encountered features results in an unsatisfiable semantic representation. Since predicted and actual features are combined, semantic anomalies like 'hippodrome' in (2) is a limiting case of the prediction-integration mechanism.

The above considerations lead to the following questions. At the empirical level one gets: (i) what neurophysiological evidence is there to support a distinction between prediction and integration/composition? (ii) given that there is a distinction between prediction and integration/composition, what type of information is predicted (atomic vs. decompositional in terms of semantic fea-

[4] Note that the pre-activated features used to predict upcoming words cannot simply be part of information about arguments, say, of verbs or common nouns. For example, 'board' in isolation does not prime (semantic features of) 'ferry' as opposed to (semantic features of) other semantically possible arguments like 'gondola' or 'airplane'.

tures)?, and (iii) predictions can be wrong; is there any empirical evidence for a stage in online semantic processing at which wrong predictions are retracted? If yes, how is this stage related to integration? These questions will be the topic of the next section where we will review electrophysiological experiments involving event-related brain potentials, in particular the N400 and two kinds of late positivity.

When implementing a predictive mechanisms in a formal semantic theory, the two principle questions are (i) in what exactly does this mechanism consist?, and (ii) where in the overall architecture of such a theory is it to be located? These questions will be the topic of the second part of this paper (see section 3). In the last part of the paper (section 4), we introduce wide-spread alternative approaches to interpret results on the N400 and P600 components (N400 as an index of semantic integration and P600 as an index of syntactic processing) and briefly discuss their shortcomings and possible implications for our theory.

2 Semantic processing online: evidence from ERPs

For semantic processing, an important event-related potential (ERP) component[5] is the N400. It is a broad, negative-going deflection that starts around 200-300 ms after a word has been presented, either auditory or visually, and peaks around 400 ms after stimulus onset. In neuroscience there is an ongoing dispute of whether the N400 reflects semantic prediction and lexical retrieval or semantic integration operations [BFH12]. In the following, we focus on the former view; section 4 critically discusses the latter approach. Thus, our approach builds on the hypothesis that the N400 is an index that allows one to examine the impact and the extent long-term memory (LTM) have on on-line semantic sentence processing. Its amplitude for a word in a given context is modulated (though not monotonic to) the word's off-line cloze probability. It was first observed in case of semantic anomalies like 'I like my coffee with cream and socks'. However, each word in a sentence elicits an N400. Furthermore, it does not even require a sentential context as shown by semantic priming tasks which involve the presentation of a semantically related or unrelated word before a target word: coffee – tea vs. chair – tea. Here 'tea' yields a larger N400 when followed after 'chair'. Note that, the N400 is *not* sensitive to negation. E.g., both 'A carrot is a fruit' and 'A carrot is not a fruit' generate more N400 activity than 'A carrot is a vegetable'.

[5] An event-related potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event. An ERP component is a portion of an ERP waveform that has a characteristic shape, timing and amplitude distribution across the scalp and a well-characterized pattern of sensitivity to experimental manipulations or neural source,[KF11,LPP08]. It is important to note that the common statement that a word does not elicit an ERP component (which will be used in this paper as well) is a simplification. It is meant that it does not trigger a brain response that significantly differs from the baseline response triggered by some control word.

2.1 Fine-grained expectations: semantic features are pre-activated

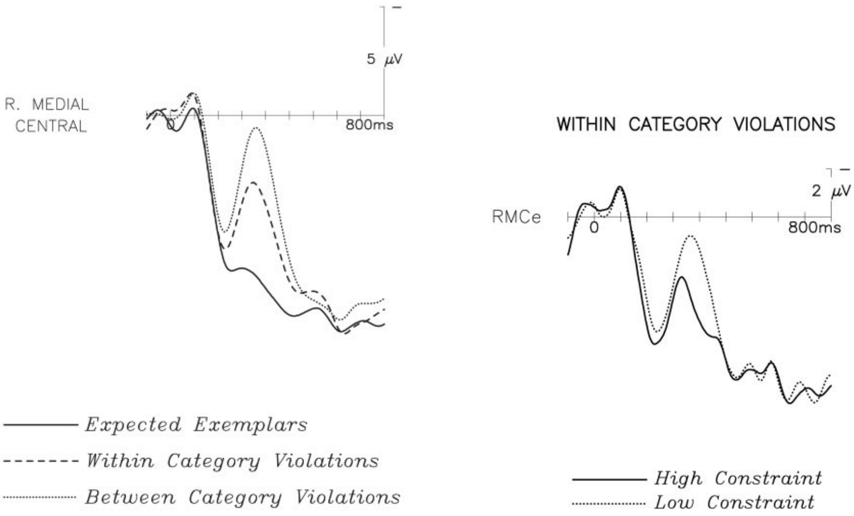
In their seminal paper [FK99], the authors investigated the following three questions w.r.t. predictions using N400 effects: (i) what type of information is predicted in a given context?, (ii) what influence do different kinds of constraining contexts have on those predictions?, and (iii) what influence do semantic relations between different target words have on the predictions? The experimental design consisted of pairs of sentences which were read by participants for comprehension. The first sentence established an expectation for a particular exemplar of a semantic category, syntactically realized by a common noun, while the second ended either (a) with this best exemplar, (b) an unexpected exemplar from the same (expected) category or (c) an unexpected exemplar from another category. An example is given in (3).

(3) They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of *palms/pines/tulips*.

Two of the three words belonged to the same taxonomic category. For example, both 'palm' and 'pine' are subtypes of the category 'tree'. The third member, 'tulip' in (3), did not belong to that category but, importantly, there was a (common) category to which all three words (or the concepts expressed by them) belong: plant. Unexpected exemplars from the same category are *within-category-violations* (WCV) whereas unexpected exemplars from another category are *between-category-violations* (BCV). Completions were ranked according to their offline cloze probability (CP, cf. footnote 3). Best completions ('palm') have highest CP. Both WCVs and BCVs had the same low CP in a given context. Additionally, sentential contexts were divided into two groups: strongly constraining and weakly constraining contexts. This distinction was defined by a median split on the CP of the best completions. For strongly constraining contexts best completions had an average value of 0.896 and in weakly constraining contexts of 0.588. WCVs and BCVs always had a CP < 0.05 across both sentential constraints. (3) is an example of a strongly constraining whereas (4) is a weakly constraining context.

(4) The gardener really impressed his wife on Valentine's day. To surprise her, he had secretly grown some *roses/tulips/palms*.

The following results were found. Overall, the N400 amplitude was significantly larger (i) for BCVs than for WCVs and (ii) for WCVs compared to best completions, i.e. one got BestComp $<$ WCV $<$ BCV (see Figure 1 for details). Strongly constraining contexts are associated with overall slightly higher, i.e. more positive, amplitudes than weakly constraining contexts, [FK99, 481]. However, there was a difference w.r.t. the factor 'constraint' for WCVs. Such violations elicited a less enhanced N400 amplitude in strongly constraining compared to weakly constraining contexts (cf. Figure 1). For both BestComp and BCVs, by contrast, there were no significant differences between the two kinds of contexts.



(a) taken from [FK99, p. 481]

(b) taken from [FK99, p. 483]

Fig. 1. (a) Comparison of N400s for BestComp (expected exemplar), WCV (within category violation), and BCV (between category violation. (b) Comparison of N400s for WCVs in strongly constraining (high constraint) and weakly constraining contexts (low constraint).

The consequences which these results have for an account of online semantic processing are the following (for details, see [FK99]). First, the information provided by the context must be rather specific. This follows from the difference in N400 amplitude between BestComp and WCVs. If only general taxonomic, say category level, information were available, members of the same category, say ‘palm’ and ‘pine’, should elicit similar brain responses. Second, the N400 is sensitive to category violations. Words that are unexpected but belong to the same category as the best completion are processed differently from unexpected ones belonging to a different category, though both words have the same (low) CP. Second, predictions/expectations come in degree and depend on the strength of the context.

According to [FK99, 489], these results constitute evidence for the view that what gets pre-activated and what is stored in LTM are semantic features of concepts expressed by words and not (discrete) atoms like ‘ferry’ or ‘palm’. The features that get activated are those associated with the best completion(s), i.e. those words having the highest CP in the given context, plus possibly features that can be inferred using world knowledge. For example, in (3) the context together with world knowledge pre-activates such features as ‘tropical’, ‘resort’, ‘adornment’, ‘tree’, and ‘evergreen’ since ‘palm’ is the best completion having the highest CP. Since three of those features equally apply to ‘pine’, its N400 amplitude though larger than that for ‘palm’ is smaller than that for ‘tulip’, for which only one feature applies. For a strongly constraining context, the number of pre-activated features is greater and therefore more constraining than the number of such features in a corresponding weakly constraining context. The more features of an upcoming word get pre-activated, the higher the probability is that even for unexpected but semantically related words (that belong to the same category) there is sufficient overlap with those features so that lexical access is facilitated. Hence, since in a strongly constraining context the number of pre-activated semantic features is greater than in a corresponding weakly constraining context, WCV should elicit a lower N400 amplitude in strongly constraining than in weakly constraining contexts, as borne out by the empirical data. Furthermore, since the overlap between pre-activated and actual semantic features is equally low for BCVs, the amplitude of the N400 should be the same for strongly constraining and weakly constraining contexts, again in line with the empirical data. Consequently, predictions/expectations should be graded and these degrees should be reflected in the corresponding amplitudes of the N400. But this is exactly what happens: BestComp $<$ WCV $<$ BCV across sentence constraint. In sum, if in a particular context a part of the semantic features representing a word A in the brain, say 'palm', is (pre-)activated, the comprehension system is better prepared to access and semantically process another word B, say 'pine', whose set of semantic features has a greater overlap to that of A than a word C, say 'tulip', for which this overlap is smaller.

2.2 The risk of pre-activation: wrong guesses

Predictions are risky because they can turn out to be wrong. E.g., if in the context of (3) 'palm' is predicted but 'pine' is eventually found, some expectations are wrong and must be deleted or retracted. Thus, there should be a stage in online semantic processing during which wrong guesses are undone. One candidate for such an operation is semantic integration. There are at least two kinds of evidence for drawing a distinction between a prediction stage in which possibly wrong features of the upcoming word are predicted and an integration stage of the semantic of the actual encountered word in which wrong features are deleted and new ones are added. First, if the N400 would be related not only to the prediction stage, but to the stage of semantic integration as well, words with the same meaning should elicit identical or very similar N400 effects, However, this is not the case as shown by the following empirical result. [DUK05] used sentence pairs like those in (5) where the sentence frame ended either with 'a' or 'an'. Since these two articles have exactly the same meaning, they should elicit the same N400 effects.

(5) The day was breezy so the boy went out to fly a/an ...

[DUK05] found a larger N400 amplitude for 'an' compared to 'a'. Since both articles have the same meaning, there should be no difference in brain response when it comes to integrating it with the semantic representation of the previous context because the semantic relation to this context must be exactly the same for both words. By contrast, if one assumes that the context preceding the article establishes a particular prediction for the most expected word 'kite', this difference can easily be explained. Since 'kite' begins with a consonant, 'a' is expected and not 'an'.

Second, there is post-N400 brain activity which is related to the semantic distinctions on which the N400 is based: late positivities. [FWODK07] considered pairs of sentences like those in (6).

(6) a. The children went outside to *play/look.* (strongly constraining context)
b. Joy was too frightened to *move/look.* (weakly constraining context)

In both kinds of context the unexpected ending, 'look' for example, had the same (low) CP.[6] In addition, the unexpected ending was not semantically related to the best completion and were considered plausible in an off-line norming task. Thus, any difference w.r.t. N400 effects could be attributed to the constraint of the sentence context. The results of the experiment showed that the N400 effects were graded by CP. The N400 amplitude was smallest for the best completion in the strongly constraining context; it was intermediate for the best completion in the weakly constraining context and highest for the unexpected completion for both kinds of constraint. However, the unexpected ending differed w.r.t. another ERP-component: a late frontal positivity between 500 and 900 ms over frontal electrode sites emerged for unexpected words in strongly constraining but not in weakly constraining contexts. The authors comment [FWODK07]: 'This processing stage thus seems to be sensitive to the greater degree of mismatch between the rich information provided by a strongly constraining sentence and an unrelated (though plausible) unexpected word, leading to the possibility of surprise and/or increased resource demands entailed by the need to override or suppress a strong prediction for a different word or concept.' This result was reproduced by [DQK14] using sentences like that in (7).

(7) For the snowman's eyes the kids used two pieces of coal. For his nose they used a *carrot/banana/groan* from the fridge.

According to [DQK14], the contexts in (7) were strongly constraining since the mean CP of the best completion, 'carrot', was 73.9%. Besides a best completion there was a semantically related and plausible continuation, 'banana', and a semantically unrelated and implausible (or impossible) continuation, 'groan' in (7). The CP for both kinds of continuation was equally low: $< 0.01\%$. In addition to the late frontal positivity the authors found an increased parietal post-N400 positivity (PNP) for unexpected and semantically implausible words. Importantly, this positivity was not exhibited by unexpected but plausible words like 'banana' in (7). Similarly, the late frontal positivity was only found for plausible but not for implausible (impossible) continuations.

Since both kinds of late positivity are not graded (in contrast to N400 effects) and apply only to one particular type of unexpected continuations, they can

[6] Cloze probabilities: 'play': 91%; 'move': 31% and 'look': 3% in both contexts.

neither be taken to simply reflect some process of plausibility evaluation nor be interpreted as a ‘mismatch’ detector.

When taken together, the results in this section provide evidence for the following picture of online semantic processing.[7] Semantic processing in the brain unfolds over several stages, [FK99,DQK14,BFH12]. The first stage is indexed by the N400 and has to do with lexical access. Semantic features of an upcoming word are activated in parallel with features of words that are currently being processed in order to access that word in LTM. The more features are already activated, the easier it is to retrieve that item from LTM. At the neural level this correlation is reflected by the amplitude of the N400: the greater the overlap with pre-activated features, and, therefore, the less features have to be additionally activated, the lower the amplitude. At this stage the item is not (yet) integrated with the semantic representation of the context. When it comes to integration, indexed by the two late positivities, what is at stake are no longer those features that are common to both the pre-activated and the actually encountered set but those feature which do not apply to the semantic representation of the word encountered. Two principle cases have to be distinguished: the target word is either of a type to which the best completion belongs or not. In the first case those features that have been pre-activated but which do not apply to the semantic representation of the word encountered have to be retracted. By contrast, in the second case, e.g. ‘groan’ in (7), a different strategy must be chosen because the semantic representation of the target word is incompatible with the semantic constraints imposed by the context.

Thus, we have arrived at the following three constraints on a formal semantic theory: (a) there must be a mechanism which combines semantic information already present in the context and world knowledge to deduce information about upcoming, but not yet presented words; (ii) the combinatory process must be sensitive to a semantic decomposition in terms of semantic features in order to account for the graded character of expectations; and (iii) there must be a separate mechanism for retracting wrong guesses made on the basis of incomplete information.

3 The formal theory: defaults and belief revision

The description at the end of the previous section suggests that online semantic processing involves some kind of nonmonotonicity. Reconsider example (3); after semantically processing the context prior to the target word at the end of the second sentence, all that is known about the concept expressed by that word is (i) the resort is supposed to look tropical and that (therefore) (ii) something is planted along the driveway. From this information conclusions about semantic features of the theme argument are drawn. Likely candidates are (a) type=plant, (b) category = tree, and (c) habitat=tropics. However, these conclusions are *defeasible*. If the upcoming word is eventually semantically recognized, the predictions made on the basis of the prior context can turn out to be false. This always

[7] Section 4 discusses alternative interpretations of the results.

happens for within-context-violations and between-context-violations. E.g., if in (3) ‘pine’ is the theme argument, ‘habitat $=$ tropics’ turns out to be false though the other predictions turn out to be true. As an effect, ‘habitat $=$ tropics’ has to be withdrawn because it is not part of the semantic representation of ‘pine’. By contrast, for the best completion ‘palm’, all information predicted before the word is encountered applies.

Nonmonotonicity will be modelled by default rules. Such rules describe the expectations of the comprehension system. Schematically, such expectations have the form $A \Rightarrow B$, with A being some piece of (factual) information provided by the context through bottom-up processing and B being the conclusions which normally follow from A. Here, ‘normally’ refers to the fact that A is all that is known about an object. The conclusions B are defeasible. For example, if in addition to $A \Rightarrow B$ one has $C \Rightarrow \neg B$ and $C \Rightarrow A$ then C is an exceptional A w.r.t. the property expressed by B. Thus, if in addition to A C is also known about the object (so that A is not only known), $\neg B$ should (normally) be true of the object. Applied to our running example of the resort which should look tropical, one has $A \doteq (i) \wedge (ii)$; $B \doteq (a) \wedge (b) \wedge (c)$ and $C = ($ sort $=$ pine $\vee$ sort $=$ tulip $)$. Thus, additional factual information can invalidate a prior inference based on less specific information. One therefore has: if both $A \Rightarrow B$ and $A \wedge C \Rightarrow \neg B$, $A \wedge C \Rightarrow \neg B$ should be used to draw the (default) conclusion $\neg B$ since one has $A \wedge C \supset A$, i.e. the antecedent of the second default rule $A \wedge C \Rightarrow \neg B$ implies that of the first one $A \Rightarrow B$. This reflects the fact that during online semantic processing conclusions drawn by more specific (less incomplete) information always overwrite conclusions drawn on the basis of less specific (more incomplete) information. What is required, therefore is an ordering on default rules which reflects this strategy. Since default conclusions can turn out to be wrong, there must be an additional mechanism of how to retract such wrong guesses. On the account just sketched, semantic processing therefore not only comprises decompositional semantic representations of items in the lexicon together with a set of recursive composition rules but, in addition, the following two components: (i) a set of default rules, which are used to draw defeasible conclusions (B) from factual information (A), and (ii) a mechanism for retracting conclusions got from applying rules in (i).

The relation to the ERP components, the N400 and the two kinds of late positivity, is the following. Default rules are correlated to the N400 and therefore to the first stage of online semantic processing. The relevant parameter is the difference between those semantic features derived after semantic recognition of the target word and those features derived prior to that recognition. This difference reflects the additional features that have to be activated. The two late positivities are correlated with those semantic features that were predicted prior to the semantic recognition of the target word but which turn out to be false and which therefore have to be retracted.

We will develop the formal theory in two steps. Building on [Bou94]. we begin by defining default rules as a conditional $\Rightarrow$ in a modal logic with a Kripke-style semantics based on a normality ordering which reflects the expectations a comprehender has for a particular constituent of a sentence in a given context. Such models are the appropriate level to reason about the whole set of defaults represented by that model. Which default conclusions can be drawn depends on the available factual information. Such reasoning is best modeled in a particular model based on a (priority) ordering on defaults. This leads to system Z, [GP92], which will be introduced in the second step.

3.1 Formal theory I: $\Rightarrow$ and CO-models

The conditional logic chosen is that of Boutilier, [Bou94]. In this theory, the conditional connective $\Rightarrow$ is not a primitive but is defined inside a modal logic using modal operators. One reason for choosing this framework is its generality. Besides default reasoning, it also allows to model belief revision. In addition, Boutilier's logic incorporates other approaches, in particular that of Pearl, [Pea90], in the sense that those logics are equivalent to fragments of Pearl's logic. This makes it possible to use either of these formalisms, depending on the context.

The basic idea underlying [Bou94] is to order situations (modeled as possible worlds in terms of valuations in a Kripke model) according to some measure of normality. This measure is represented by an accessibility relation $\geq_N$ on worlds. One has $w \geq_N v$ iff v is at least as normal as w. $w >_N v$ holds if v is strictly more normal than w, that is if $w \geq_N v$ and not $v \geq_N w$. The relation $\geq_N$ is required to be (i) transitive and (ii) totally connected from which together reflexivity follows: (i) $\forall uvw : u \geq_N w \wedge w \geq_N v \supset u \geq_N v$, and (ii) $\forall wv : w \geq_N v \vee v \geq_N w$. Models in which (i) and (ii) hold consist of totally ordered clusters of worlds, where a cluster is any maximal set of worlds s.t. $w \geq_N v$ for each w, v in this set, i.e. the elements of a cluster are all equally normal and the cluster is maximal w.r.t. this condition. If the set of worlds is finite, this chain of clusters has both a minimal and a maximal element. Furthermore, this ordering determines a normality ranking for each cluster and, therefore, for each world in W.[8]

Next, the language L_{Frame} is defined. As was shown in the first section, the information predicted is rather specific. We will therefore use a frame-based approach [Pet07]. Frames are recursive rooted attribute-value structures.[9] A modal language for talking about such structures is given by a set $\{P_\sigma\}_{\sigma \in \Sigma}$ of sort symbols ($\Sigma = \{\mathit{tree}, \mathit{palm}, \ldots\}$) and a set $\{Attr_{at}\}_{at \in ATTR}$ of attribute

[8] The ordering $\geq_N$ depends both on the kind of context and the comprehender. The dependency on the context corresponds to the distinction between strongly constraining and weakly constraining contexts. In a strongly constraining context there are more expectations than in a weakly constraining context. The dependency on a comprehender is illustrated by the following example concerning the moral value system of a comprehender. [BHN$^+$09] presented examples like 'I think euthanasia is an acceptable course of action' to members of a relatively strict Dutch Christian party and to non-Christian respondents with sufficiently contrasting moral value systems. The result was that for both groups there was an enhanced N400 though it was larger for members of the strict Dutch Christian party.

[9] Note that [Pet07] allows unrooted frames as well, but such frames are of no interest for our purpose.

symbols ($ATTR = \{habitat, look, \ldots\}$). Elements of $\{P_\sigma\}_{\sigma \in \Sigma}$ are interpreted as unary relations and elements of $\{Attr_{at}\}_{at \in ATTR}$ as binary relations on a set of nodes. Formulae are of the form $at_1 : at_2 : \ldots at_n = \sigma$, expressing that the value at the end of the sequence of attributes $at_1 : at_2 \ldots at_n$ is of sort σ. They therefore express properties of nodes, as can be seen by looking at the standard translation of such a formula in first-order logic: $\lambda x \exists y_1 \ldots \exists y_n . at_1^*(x, y_1) \wedge \ldots \wedge at_n^*(y_{n-1}, y_n) \wedge \sigma^*(y_n)$ (see [PO14] for details). By interpreting such formulae at the root of a frame, a frame can be described by what is true at its root. On this perspective, frames can be taken as points (possible worlds) in a model. Formulae of the form $at_1 : at_2 \ldots at_n = \sigma$ are then atomic propositions in the language L_{Frame}. In addition, L_{Frame} has three modal operators $\Box$, $\overleftarrow{\Box}$ and $\Box_>$. While $\Box A$ refers to all accessible (i.e. equally or more normal) worlds in the ordering $\geq_N$, $\overleftarrow{\Box}\, A$ means that A is true at all inaccessible worlds, i.e. at all worlds which are strictly less normal than the world at which $\overleftarrow{\Box}$ is evaluated. $\Box_>$ is the strict variant of $\Box$. Models for L_{Frame} are defined below.

Definition 1 (A CO-model; [Bou94, 101]) *A CO-model is a triple $\langle W, \geq_N, V\rangle$ s.t. (i) W is a non-empty, finite set of worlds, (ii) $\geq_N$ is a binary relation on W that is transitive and totally connected and (iii) V is a valuation function for the atomic formulas in L_{Frame}.*

Truth of a formula is defined as follows.

Definition 2 *Let $M = \langle W, \geq_N, V\rangle$ be a CO-model with $w \in W$. The truth of a formula A at w in M is defined inductively by*

(i) $M \models_w A$ *iff* $w \in V(A)$ *for atomic sentence* A.
(ii) $M \models_w A \supset B$ *iff* $M \models_w B$ *or not* $M \models_w A$.
(iii) $M \models_w \neg A$ *iff not* $M \models_w A$.
(iv) $M \models_w \Box A$ *iff for each* v *s.t.* $w \geq_N v : M \models_v A$.
(v) $M \models_w \overleftarrow{\Box}\, A$ *iff for each* v *s.t.* $w \not\geq_N v : M \models_v A$.
(vi) $M \models_w \Box_> A$ *iff for each* v *s.t.* $w >_N v : M \models_v A$.

In terms of $\Box$ and $\overleftarrow{\Box}$ the following modal operators are defined.

Definition 3 (Defined modal operators)

1. $\Diamond A \equiv_{df} \neg \Box \neg A$.
2. $\overleftarrow{\Diamond}\, A \equiv_{df} \neg \overleftarrow{\Box} \neg A$.
3. $\overleftrightarrow{\Box}\, A \equiv_{df} \Box A \wedge \overleftarrow{\Box}\, A$.
4. $\overleftrightarrow{\Diamond}\, A \equiv_{df} \Diamond A \wedge \overleftarrow{\Diamond}\, A$.

One has: $\Diamond A$ is true at $w \in W$ iff A is true at some equally or more normal world v; similarly, $\overleftarrow{\Diamond}\, A$ holds at w just in case A holds at some strictly less normal world v; $\overleftrightarrow{\Box}\, A$ holds at a world w iff A is true at each world $w \in W$; $\overleftrightarrow{\Diamond}\, A$ is true at w iff A is true somewhere in the model, i.e. if there is a world $v \in W$ at which A is true. The conditional $\Rightarrow$ is defined in Definition 4.

Definition 4 ([Bou94, 104]) $A \Rightarrow B \equiv_{df} \overset{\leftrightarrow}{\Box} \neg A \vee \overset{\leftrightarrow}{\Diamond} (A \wedge \Box(A \supset B))$.

According to Definition 4, $A \Rightarrow B$ is true at a world w just in case either A is false at every world in the chain of worlds, i.e. the conditional is satisfied vacuously, or at the most normal A-worlds $(A \supset B)$ holds. The truth of $A \Rightarrow B$ is independent of a particular possible world. If $A \Rightarrow B$ holds at some w, then it holds at all $v \in W$. This follows from the fact that the disjuncts in the definition of $\Rightarrow$ are modally decorated by $\overset{\leftrightarrow}{\Box}$ and $\overset{\leftrightarrow}{\Diamond}$, respectively. As a consequence, the truth of $A \Rightarrow B$ only depends on the complete ordering of worlds.

A CO-model represents the set of default rules Δ_D of a comprehender w.r.t. an argument (or a constituent) of a sentence in a given context. Together with factual information A got from bottom-up processing of the prior context (and, possibly, world knowledge), default rules $A \Rightarrow B$ are used to (defeasibly) infer B. More generally, one has: the local epistemic state of a comprehender w.r.t. an upcoming word is a quadruple $ES = \langle \Gamma, \Gamma^*, \Delta_D, \Delta_E \rangle$. Δ_D is a set of defaults of the form $A \Rightarrow B$ and Δ_E is the set of expectation rules given by the corresponding material conditionals $A \supset B$.[10] Γ is the set of factual information about the word. Before the word is semantically recognized it contains information got from the context. Upon recognition of the word, sortal information, e.g. *sort=palm* is added. Γ^* is a set of default conclusions pertaining to the target word. They are inferred using Γ and Δ_E.

The reason for distinguishing Γ and Γ^* is directly related to the way semantic information is used in default rules $A \Rightarrow B$. The antecedent contains factual information from bottom-up semantic processing. This information is stored in Γ. By contrast, the information B in the consequent of a default rule is used to build up a partial semantic representation of an upcoming word. Since this information is in general defeasible (the problem of 'wrong guesses'), it is not directly integrated with the factual information stored in Γ but stored separately in Γ^*. This reflects the distinction between *lexical access* (first stage of semantic processing) and *integration* (second stage of semantic processing). During semantic processing, Γ and Γ^* are constantly updated whereas both Δ_D and Δ_E remain fixed.

3.2 Defaults and online semantic processing

Next we will apply CO-models to online semantic processing. As our running example we will take (3), repeated below for convenience.

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

After processing (8) up to the final world, the comprehender has got the following factual information which is relevant for drawing default conclusions about the object planted.

[10] The reason for distinguishing Δ_D and Δ_E will become clear if a ranking on the set Δ_D of default rules using System Z is defined. See below for details.

(9) a. $\textit{resort:look}=\textit{tropical}.$
b. $\textit{resort:driveway:adornment}=\top.$

Let this information be A_0. This information is related to the following default rules.

(10) a. $A_0 \Rightarrow \textit{resort:driveway:adornment:type}=\textit{plant}.$
b. $A_0 \Rightarrow \textit{resort:driveway:adornment:category}=\textit{tree}.$
c. $A_0 \Rightarrow \textit{resort:driveway:adornment:sort:habitat}=\textit{tropics}.$

When taken together, one gets default rule r_0 in (11).

(11) $r_0 :\ A_0 \Rightarrow$
$\textit{resort:driveway:adornment:type}=\textit{plant} \wedge$
$\textit{resort:driveway:adornment:category}=\textit{tree} \wedge$
$\textit{resort:driveway:adornment:sort:habitat}=\textit{tropics}.$

The material conditional r_0^* corresponding to r_0 is (12).

(12) $r_0^* :\ A_0 \supset$
$\textit{resort:driveway:adornment:type}=\textit{plant} \wedge$
$\textit{resort:driveway:adornment:category}=\textit{tree} \wedge$
$\textit{resort:driveway:adornment:sort:habitat}=\textit{tropics}.$

What happens if the upcoming word is eventually encountered and semantically recognized? In our frame theory, the information provided by a common noun like ‘palm’ is taken as sortal information. In (8), this is the value of the *sort*-attribute. Thus, if ‘palm’ is semantically recognized

$\textit{resort:driveway:adornment:sort}=\textit{palm}$

is added to Γ. The default rule corresponding to this information is r_1.

(13) r_1: $A_0 \wedge \textit{resort:driveway:adornment:sort}=\textit{palm} \Rightarrow$
$\textit{resort:driveway:adornment:type}=\textit{plant} \wedge$
$\textit{resort:driveway:adornment:category}=\textit{tree} \wedge$
$\textit{resort:driveway:adornment:sort:habitat}=\textit{tropics}.$

Rule r_1 differs from r_0 in one respect. Its antecedent is more specific than that of r_0 ($A_1 \supset A_0$). This reflects the fact that r_0 is used in a situation of incomplete information, i.e. the upcoming word has not yet been semantically recognized whereas r_1 is used after that recognition has taken place. The consequents are the same because ‘palm’ is the best completion and therefore all predicted properties apply to the word encountered. The general pattern between these two default rules is given in (14).

(14) $r_0 :\ A \Rightarrow B.$
$r_1 :\ A \wedge C \Rightarrow B.$

This pattern can be taken as showing that encountering the best completion amounts to a confirmation of the expectations drawn when this word is not yet encountered.[11] The situation is different if instead of the best completion a within-context-violation like ‘pine’ is found. Similar to the case of ‘palm’, new sortal information is added to the factual information,

Γ: $\mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{pine}$.

One also has that the antecedent of the corresponding default rule is more specific than that of r_0. But in this case the two consequents are logically incompatible because B_0 contains $\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat}=\mathit{tropics}$ whereas B_2 contains $\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat}=\mathit{moderate}$.

(15) r_2: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{pine} \Rightarrow$
$\mathit{resort{:}driveway{:}adornment{:}type}=\mathit{plant} \wedge$
$\mathit{resort{:}driveway{:}adornment{:}category}=\mathit{tree} \wedge$
$\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat}=\mathit{moderate}$.

The general relation between the two default rules is given in (16).

(16) $r_0 :\ A \Rightarrow B.$
$r_2 :\ A \wedge C \Rightarrow \neg B.$

The case for ‘tulip’ should by now pose no problems. The default rule is r_3.

(17) r_3: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{tulip} \Rightarrow$
$\mathit{resort{:}driveway{:}adornment{:}type}=\mathit{plant} \wedge$

[11] According to rule r_0, an expectation w.r.t. to the theme argument of ‘plant’ does not include sortal information. Thus, there is no bias towards any tropical tree in the context of A_0. For example, both ‘palm’ and ‘eucalyptus’ are equally expected. However, if ‘palm’ is the best completion one may argue that this information is already activated prior to the encounter of the argument. Thus, rule r_0 seems to apply to weakly constraining and not to strongly constraining contexts. However, if sortal information is part of the consequent of the default rule, alternatives (‘eucalyptus’) to the best completion (‘palm’) are excluded. E.g., rule r_0 becomes r_{00}.

(i) r_{00} $A_0 \Rightarrow B_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{palm}$.

Using r_{00}, r_1 becomes redundant because upon encountering ‘palm’ no new information needs to be added. Rule r_1 is replaced by the following rule for the sort ‘eucalyptus’.

(ii) r_1: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{eucalyptus} \Rightarrow$
$B_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort}=\mathit{eucalyptus}$.

An open empirical question is the relation between N400 effects both in strongly constraining and weakly constraining contexts for ‘palm’ and ‘eucalyptus’, i.e. two concepts that are of the same type, here ‘plant’, but also of the same category. here ‘tree’, and that both fulfill the conditions specified in the consequent of rule r_0.

$\mathit{resort{:}driveway{:}adornment{:}category{=}flower} \wedge$
$\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat{=}moderate}.$

Similar to the case of 'pine', the consequent is logically incompatible with that of r_0 (and also with that of r_2). In contrast to 'pine', there are two conjuncts which are logically incompatible. Besides the one specifying the value of the HABITAT-attribute, this also holds for the value of the SORT-attribute.[12]

A drawback of the rules r_1–r_3 is that they contain redundant information. This is the case whenever they contain information that is also specified in the rule r_0. This information will not be retracted even when a non-best completion is encountered. An alternative is to only specify that information which is incompatible with information given by r_0. Applied to the processing level, this means that once a feature is activated it need not be activated a second time. At the formal level, one uses the following property of formulae.

Definition 5 (Downward closed property) *A formula A is downward closed iff* $\overset{\leftrightarrow}{\Box}(A \supset \Box_{>} A)$.

According to this definition, a formula is downward closed if its truth at a world w implies that it holds at all strictly more normal worlds. The revised rules $r'_1 - r'_3$ are given in (18).

(18) r'_1: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort{=}palm} \Rightarrow \mathit{true}.$
r'_2: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort{=}pine} \Rightarrow$
$\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat{=}moderate}.$
r'_3: $A_0 \wedge \mathit{resort{:}driveway{:}adornment{:}sort{=}tulip} \Rightarrow$
$\mathit{resort{:}driveway{:}adornment{:}category{=}flower} \wedge$
$\mathit{resort{:}driveway{:}adornment{:}sort{:}habitat{=}moderate}.$

A possible model for the default rules is given in Figure 2. This model is based on a knowledge base corresponding to our running example: the objects planted are either palms, pines or trees and there are no 'abnormal' instances of those sorts.[13]

[12] One may argue that rules $r_1 - r_3$ are strict and not defeasible. For example, a palm is a tree and not a flower. However, in the present context we are interested in the way a comprehender uses information, both top-down and bottom-up, to build a semantic representation of a constituent. What matters, therefore, is the relation between the various rules he uses (the priority ordering) and not the status of an individual rule as defeasible or strict. For example, rule r_2 has a higher priority than rules r_0 and r_1 because it describes a situation which is assumed to be less normal. In addition, not all conjuncts in the consequent of a rule are non-defeasible, given the antecedent. For example, the tropics are only normally the habitat of palms, but they grow in moderate habitats as well (e.g., in botanical gardens in Europe).

[13] These restrictions are due to the fact that we do not have any information about the way, say, orchids (tropical flowers) or palms whose habitat are not the tropics are semantically processed online. Additional experimental data is needed to tackle this question.

Cluster	**0** (palm)	**1** (pine)	**2** (tulip)
tropics	true	false	false
tree	true	true	false
flower	false	false	true
plant	true	true	true

Fig. 2. Possible model for the running example

Cluster	$\Box A \wedge \overset{\leftarrow}{\Box} \neg A$
0	$A \doteq \mathit{habitat{=}tropcis}$
1	$A \doteq \mathit{category{=}tree}$
2	$A \doteq \mathit{type{=}plant}$

Fig. 3. Relation between clusters and properties of objects adorning the driveway

In L_{Frame}, the four clusters can be formally characterized as follows. To begin, note that the formula $\Box A \wedge \overset{\leftarrow}{\Box} \neg A$ holds at a world w_0 if A is true at all equally or more normal worlds w_1 whereas at all worlds w_2 which are strictly less normal A is false. This formula can therefore be seen as expressing a kind of 'frontier'. All worlds above the frontier satisfy A whereas all worlds below it fail to satisfy it. The relation between this formula, properties of the objects adorning the driveway and clusters are shown in Figure 3. Thus, cluster 0 is a frontier for the property *habitat=tropcis* (for ease of readability, only the last attribute of a chain of attributes is displayed) whereas clusters 1 and 2 are frontiers for the properties *category=tree* and *type=plant*, respectively. This correlation between clusters and properties shows that of the three properties assumed in a most normal situation, *habitat=tropics* is the least entrenched one or the first to be given up. Similarly, *type=plant* is the most entrenched one whereas *category=tree* has a position intermediate between those two properties. Intuitively, one can say that 'tropics'-worlds only see 'tropics'-world and similarly for 'tree'- and 'plant'-worlds. The difference shows up if one looks backwards. 'tree'-worlds are either seen by 'non-tropics'-worlds or if in the same non-minimal cluster, 'tree'-worlds are always 'non-tropics'-worlds: $\overset{\leftrightarrow}{\Box}(\mathit{tree} \supset (\mathit{tree} \wedge \neg \mathit{tropics}) \vee \overset{\leftarrow}{\Box} \neg \mathit{tropics})$. Thus, 'tropics'-worlds are more normal than 'tree'-worlds. Furthermore, one has $\overset{\leftrightarrow}{\Box}(\mathit{flower} \supset \Diamond \Box \mathit{tree})$: 'flower'-worlds are no more normal than 'tree'-worlds. Finally, one has $\overset{\leftrightarrow}{\Box}(\mathit{plant} \supset (\mathit{tree} \vee \mathit{flower}))$. The above properties are global in the sense that their truth is independent of a particular world.

General CO-models are appropriate for specifying global properties of the local epistemic state of a comprehender w.r.t. an upcoming word. If a comprehender uses a CO-model to draw conclusions, it is more convenient to use a particular CO-model which is based on a priority ordering on default rules.

3.3 Formal theory II: Defining an ordering on defaults

An ordering on default rules can be defined using procedure Z, [GP92]. Defeasible rules can be verified, falsified or satisfied at a world w.

Definition 6 (Verifying, falsifying and satisfying a default rule) *A possible world w in a model M* verifies *a conditional $A \Rightarrow B$ iff $M \models_w A \wedge B$; it* falsifies *$A \Rightarrow B$ iff $M \models_w A \wedge \neg B$, and it satisfies $A \Rightarrow B$ iff $M \models_w A \supset B$.*

The derivation of a Z-ordering of default rules is based on the notion of *toleration*, Definition 7.

Definition 7 (Toleration) *Δ_D is said to* tolerate *a default $A \Rightarrow B$ iff there is a world w that verifies $A \Rightarrow B$ and falsifies no rule in Δ_D, i.e.*

(19) $$A \wedge B \wedge \textstyle\bigwedge_{r_j \in \Delta_D} A_j \supset B_j.$$

Toleration is used to define a natural ordering on a set of defaults by partitioning this set. The procedure for finding this partition works as follows. Let Δ be the set of defaults. In a first step all rules in Δ which are tolerated by all other rules are in Δ_0. Next, the set $\Delta' = \Delta - \Delta_0$ is considered. All rules in Δ' which are tolerated by all other rules in Δ' are in Δ_1. Next, the set $\Delta'' = \Delta' - \Delta_1$ is considered. Continuing in this way, yields a partition $\Delta_0, \Delta_1, \ldots, \Delta_n$ of Δ (provided Δ is consistent). This procedure is defined inductively in Definition 8 where $\Gamma(\Delta)$ is the set of defaults in Δ which are tolerated by Δ.

Definition 8 (Partition of a set of defaults) *$\Delta_0 = \Gamma(\Delta)$ and $\Delta_{\tau+1} = \Gamma(\Delta - (\bigcup_{\sigma \leq \tau} \Delta_\sigma))$*

Given this partition of Δ, the rank of a default $A \Rightarrow B \in \Delta$ is defined by $Z(A \Rightarrow B) = \tau$ iff $A \Rightarrow B \in \Delta_\tau$. The intuition is that lower ranked defaults are more general and have a lower priority.

Next, the ranking of a world w is defined. The rank of a world w is the smallest integer τ s.t. all defaults having a rank higher or equal to τ are not falsified by w. This condition is expressed by: w satisfies $\bigcup_{\sigma \geq \tau} \Delta_\sigma$ or, equivalently by $Z(w) = min\{\tau : M \models_w A \supset B$ for all $r \in \Delta$ and $Z(r) \geq \tau\}$. The intuition is that lower ranked worlds are more normal. Thus, the Z-ranking on worlds determines a unique preferred structure Z_T.

The rank of a (non default) formula A is defined as follows.

(20) $$\kappa^z(A) = min\{i \,|\, A \wedge \textstyle\bigwedge_{j:Z(r_j) \geq i} A_j \supset B_j \textit{ is satisfiable}\}.$$

Using this ranking on formulae, a formula B is said to be Z-entailed by a formula A iff the worlds in which A and B hold are strictly lower ranked than the worlds in which A and $\neg B$ hold, that is if the rank of $A \wedge B$ is strictly lower than the rank of $A \wedge \neg B$.

Definition 9 (Z-entailment) *A formula B is Z-entailed by a formula A w.r.t. Δ, written $A \vdash_Z B$, iff $\kappa^z(A \wedge B) < \kappa^z(A \wedge \neg B)$.*

(19) and (20) can be used to construct a theory $\mathit{Th}(A)$ which characterizes precisely the set of conclusions B that defeasibly follow from factual information A, given a set $\varDelta_D$ of default rules: $A \vdash_Z B$ iff $\mathit{Th}(A) \supset B$.

(21) $\mathit{Th}(A) = A \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A)} A_i \supset B_i$.

In our application, A is always factual information about an upcoming word (or an argument) got by bottom-up processing and stored in $\varGamma$. $\varDelta_D$ (or $\varDelta_E$) is a set of default rules (expectations) which pertain to this argument. In our running example, this is the theme argument of the verb 'plant' in a given context. The $A_i \supset B_i$ are elements of $\varDelta_E$, i.e. material counterparts of default rules in $\varDelta_D$. The elements of $\varGamma^*$ are those B_i which follow from $\mathit{Th}(A)$, i.e. from A and the $A_i \supset B_i$ with $A \supset A_i$.

3.4 Drawing default conclusions from factual information

Let us next apply system Z to our running example. We first construct a Z-ranking on $\varDelta_D = \{r_0, r_1, r_2, r_3\}$. Rules r_1 and r_0 are tolerated by all the other rules. The following valuation verifies both rules:

> *resort:look=tropical* $\wedge$ *resort:driveway:adornment:*$\top$
> $\wedge$ *resort:driveway:adornment:type=plant* $\wedge$
> *resort:driveway:adornment:category=tree* $\wedge$
> *resort:driveway:adornment:sort:habitat=tropics* $\wedge$
> *resort:driveway:adornment:sort=palm.*

Furthermore, one sets $\neg$*resort:driveway:adornment:sort=X* for $X \in \{\mathit{pine}, \mathit{tulip}\}$. Since the antecedents of the rules r_2 and r_3 are pairwise logically incompatible, each rule tolerates the others. For example, verifying r_2 requires

> *resort:driveway:adornment:sort=pine.*

Setting $\neg$*resort:driveway:adornment:sort=tulip* satisfies r_3. Here it is assumed that one has e.g. *tree* $\supset \neg$*flower* . Therefore, for $j \neq k$ with $j, k \in \{2, 3\}$ we get that if a world verifies $A_j \Rightarrow B_j$, it satisfies $A_k \Rightarrow B_k$ because A_k is false at that world. The Z-ranking on rules is $\varDelta_0 = \{r_0, r_1\}$ and $\varDelta_1 = \{r_2, r_3\}$.

As long as no factual information about the theme is given, one has $A = \mathit{true}$. No conclusions using the set of expectations $\varDelta_E$ can be drawn. Furthermore, $\varGamma = \{\mathit{true}\}$, $\varDelta_D = \{r_0, r_1, r_2, r_3\}$, $\varDelta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$ and $\varGamma_0^* = \emptyset$. After processing the prior context, one has $A = A_0$ and $\varGamma = \{A_0\}$. Since $\kappa^Z(A_0) = 0$, one gets $\mathit{Th}(A_0) = A_0 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_0)=0} A_i \supset B_i$. Thus, $\varDelta_D = \{r_0, r_1, r_2, r_3\}$ and $\varDelta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$. The set of defeasible consequences $\varGamma_0^*$ is deduced from $A = A_0$ and $A_0 \supset B_0$ yielding $\varGamma_0^* = \{B_0\}$. If 'palm' is encountered, the sortal information *sort=palm* is added to $\varGamma$ so that $A = A_1$. Since $\kappa^Z(A_1) = 0$, one has $\varDelta_D = \{r_0, r_1, r_2, r_3\}$ and $\varDelta_E = \{r_0^*, r_1^*, r_2^*, r_3^*\}$. The set of defeasible consequences is got from A_0, A_1, and $A_0 \supset B_0$ and $A_1 \supset B_1$, which yields $\varGamma_1^* = \{B_1\}$ since $B_1 \supset B_0$. If instead of r_1, r_1' is used no new (defeasible) information is added to $\varGamma^*$.

If a within-context-violation or a between-context-violation is encountered, the new sortal information is *sort=pine* or *sort=tulip* in our running example. It is added to $\mathit{\Gamma}$, yielding $A = A_2$ ('pine') or $A = A_3$ ('tulip'). In contrast to A_0 or A_1, one has $\kappa^Z(A_2) = \kappa^Z(A_3) = 1$ so that $\mathit{Th(A_2)} = A_2 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_2)=1} A_i \supset B_i$ and $\mathit{Th(A_3)} = A_3 \wedge \bigwedge_{i:Z(r_i)\geq\kappa^z(A_3)=1} A_i \supset B_i$. This means that the situation is not described as most normal. As a result, r_0 and r_1 can no longer be used. One rather gets $\mathit{\Delta}_D = \{r_2, r_3\}$ and $\mathit{\Delta}_E = \{r_2^*, r_3^*\}$.

For A_2 ($\doteq$ *sort* $=$ *pine*), the conclusions one gets are given by A_2, $A_2 \supset B_2$, yielding B_2. Using r_2' instead of r_2^*, one has $B_2 = \{\mathit{habitat{=}moderate}\}$, i.e. $\mathit{\Gamma}_1^* = \{\mathit{habitat{=}moderate}\}$. For the BCV 'tulip', the situation is similar. Conclusions are got from A_3 and $A_3 \supset B_3$, yielding B_3. Using r_3' instead of r_3^*, the new derived information is *habitat=moderate* and *category=flower*, i.e. $\mathit{\Gamma}_1^* = \{\mathit{habitat{=}moderate}, \mathit{category{=}flower}\}$. Both for A_2 and A_3, it is not possible to directly add B_2 or B_3 to $\mathit{\Gamma}^*$, i.e. to use $\mathit{\Gamma}_0^* \cup \mathit{\Gamma}_1^*$. This would result in an unsatisfiable set because one would have both *habitat=tropics* (from the previous application of rule r_0 prior to the semantic recognition of the theme) and *habitat=moderate* from applying r_2^* or r_3^*. In addition r_3^* yields *category=flower* which conflicts with *category=tree*, again got from applying r_0 prior to encountering the theme argument. Despite the fact that $\mathit{\Gamma}_0^*$ (got from applying r_0^*) and $\mathit{\Gamma}_1^*$ (the information got from applying r_2^* or r_3^*) are logically incompatible, their union contains all semantic features necessary for building up a semantic representation of the theme argument.

Let us take stock and compare a best completion, a within-context-violation and a between-context-violation. One has: (a) in each case sortal information is added to the default conclusions got prior to encountering the argument, (b) they differ w.r.t. the set $\mathit{\Gamma}_1^* - \mathit{\Gamma}_0^*$, and (c) they differ w.r.t. the set $\mathit{\Gamma}_0^* - \mathit{\Gamma}_1^*$. The set $\mathit{\Gamma}_1^* - \mathit{\Gamma}_0^*$ is the set of semantic features that have to be activated in addition to those that were activated prior to the semantic recognition of the target word. By contrast, the set $\mathit{\Gamma}_0^* - \mathit{\Gamma}_1^*$ (using the rules r_i and not the rules r_i') is the set of semantic features that have to be retracted because they are 'wrong guesses'. Now consider the two hypotheses in (22).

(22) (i) The set $\mathit{\Gamma}_1^* - \mathit{\Gamma}_0^*$, i.e. the set of additional features to be activated, is related to the N400 effect, i.e. it is related to the first stage of online semantic processing: semantic access.

(ii) Predicting semantic features of an upcoming word can lead to wrong guesses. Those wrong guesses must be eliminated before the semantic representation of the target word can be added to the representation of the prior context. The set $\mathit{\Gamma}_0^* - \mathit{\Gamma}_1^*$, containing those wrong guesses, is related to the two late positivities and therefore to the second stage of online semantic processing.

In the introduction it was argued that online semantic processing can be split in (at least) two separate stages: lexical semantic access, indexed by the N400, and semantic integration, indexed by two late positivities. The former, lexical access, is based on predictions which are made prior to encountering the target word,

word encountered	$\lvert\Gamma^*_{r_i} - \Gamma^*_{r_0}\rvert$	N400 amplitude
palm (best completion)	0	base line
pine (within-category violation)	1	a > base line
tulip (between-category violation)	2	b > a

Fig. 4. Default rules and N400 effects

and, therefore, on the basis of incomplete information. Such predictions are risky because they can turn out to be wrong. On the account presented in this paper, wrong guesses are directly related to the two stages of online semantic processing. A wrong guess activates less semantic features of the actual target word; thus, lexical access is aggravated. Accessing additional semantic features comes with a computational cost because information needs to be retrieved from LTM. This cost is reflected in an enhanced amplitude of the N400. This is only one side of the coin. The other is, of course, that a wrong guess has to be retracted. This follows from the fact that predictions, be they based on incomplete information or on information after the word is encountered, are related to accessing the associated features in LTM. Thus, once a semantic feature has been activated using rule r_0, it has to be retracted if it turns out to be wrong after the target has been semantically recognized and before the target is integrated into the prior context. This operation is related to the second stage, the integration stage. As a result, integration becomes a two stage process: first retracting wrong guesses and only then integrating the semantic representation of the target with the representation of the prior context. The above correlations will be explained in more detail in the following sections.

3.5 The N400 and default reasoning

We hypothesize the following correlation between the N400 effect and default reasoning.

(23) *Correlation N400 – default reasoning:*
The N400 effect is monotonic to the difference between semantic features got after semantic recognition of the target word and prior to its semantic recognition.

According to (23), the N400 effect is correlated to the difference between the pre-activated features $\Gamma^*_{r_0}$ if only rule r_0 is used, representing the most normal expectations, and those features contained in the consequent of the rule used after the upcoming word is eventually being semantically recognized. One calculates the cardinality of $\Gamma^*_{r_i} - \Gamma^*_{r_0}$. The greater this cardinality, the greater the N400 effect. Thus, the N400 is a measure of the cost of activating additional semantic features after recognition of the target word. For our running example, this correlation is shown in Figure 4.

If ‘palm’ is encountered, rules r_0 and r_1 apply. As shown above, no new features need to be activated so that all features already pre-activated become part of the frame-representation of the concept expressed by this word. If ‘pine’ is encountered, neither rule r_0 nor rule r_1 apply. Instead rule r_2 is used. In this case only one feature does not apply: *habitat=tropics* so that one new feature *habitat:moderate* of the concept expressed by ‘pine’ must be activated. For ‘tulip’, the situation is similar. The difference is that even fewer pre-activated features apply: *type=plant*. Therefore, more additional features have to be activated: *category=flower* and *habitat=moderate*.

The above criterion for the amplitude of the N400 can be refined in the following way. Instead of just counting the number of attributes, one considers in addition the degree of entrenchment of an attribute. For example, the attribute ‘category’, with its value ‘plant’, is more entrenched than the attribute ‘type’, with values ‘tree’ or ‘flower’. Formally, such distinctions can be made in an extension of system Z, system Z*, [GP91]. In Z*, a default rule is of the form $A \overset{\delta}{\Rightarrow} B$. Intuitively, δ is a measure of strength or the degree of surprise of finding the default rule violated. Applied to the above example, one gets: the value of δ for *type=plant* is greater than that for *category=tree*.

3.6 Late positivities and belief revision

Frontal late positivity One difference between a best completion on the one hand and a within-context-violation and a between-context-violation on the other is the fact that for the latter but not for the former there are wrong guesses. At the formal level, this corresponds to the distinction between $\Gamma_0^* - \Gamma_1^*$ being empty or not. If this set is empty, the default conclusions drawn before the target word is encountered are completely confirmed. Formally, this process is an addition. First, A, the sortal information, is added to Γ^* and next $\Gamma^* \cup \{A\}$ is added to Γ.

(24) a. $\Gamma_{i+1}^* = \Gamma_i^* \cup \{A\}$.
b. $\Gamma_{i+1} = \Gamma_i \cup \Gamma_{i+1}^*$.

If a non-best completion is encountered, processing is more involved. This is a simple consequence of the fact that the comprehender knows that the situation described is not most normal and that therefore at least some of his expectations are not satisfied. First, default rule r_0 can no longer be used because the theory w.r.t. the target word has changed. Instead of $Th(A_0)$, $Th(A_i)$ with $2 \leq i \leq 3$ has to be used. Second, $Th(A_0)$ and $Th(A_i)$ are incompatible. Using (21), this is the case for the information B_i contained in the consequent of rule r_i. For example, if ‘pine’ is encountered, one gets *habitat=moderate*, which is incompatible with *habitat=tropics* got from applying r_0 during the first stage. Let this information, i.e. *habitat=moderate*, be A. One then has $\Gamma^* \models \neg A$. Therefore, it is not possible to simply add A to Γ^* because this would result in a set which is not satisfiable. Rather, one first has to retract $\neg A$ from Γ^*. Only after this has been done, the addition operation given in (24) can be applied. Formally, this amounts to a revision operation in terms of the Levi-identity.

(25) Levi-identity:$KB \overset{*}{-} A = (KB \overset{\cdot}{-} \neg A) + A$.

Revising a knowledge base KB with A amounts to first making KB consistent with A by removing $\neg A$ from KB and then adding A to the resulting KB from which $\neg A$ has been retracted. For a best completion, the retraction step does not apply because there is no new default information which is incompatible with information got during the first stage. As an effect, revising reduces to a simple addition. In the present context, KB is always Γ^*, i.e. the set of default conclusions got by applying r_0. A is the conjunction of the literals in the consequent of rule r_i, i.e. the conjunction of those literals which differ in the value assigned to an attribute from those in the consequent of r_0. Thus, the retraction operation is always applied to Γ^* and therefore to defeasible conclusions. This reflects the fact that defeasible information, i.e. information got from top-down processing using default rules, is always less entrenched than information got by bottom-up processing.

We hypothesize the following relation between the frontal late positivity and the formal process described above.

(26) *Correlation frontal late positivity – belief revision:*
A frontal late positivity is triggered whenever $\Gamma^* \overset{*}{-} A$ is a *proper* revision, i.e. if there is a non-empty retraction operation. In this case default conclusions drawn before the target word is encountered have to be retracted.

Parietal late positivity As was shown in the previous section, the revision of Γ_i^* by the new information got from processing the target word is successful, not only for the best completion but also for a within-context-violation or a between-context-violation. This follows from the fact that both kinds of violation satisfy the minimal appropriateness condition imposed on the theme argument of the verb ‘plant’, namely the type of the object must be ‘plant’. At the level of CO-models, this is expressed by the integrity constraint, $\overset{\leftrightarrow}{\Box}$ $(\mathit{type}{=}\mathit{plant})$. The effect of this constraint is that any attempt to update with information which does not satisfy this constraint, say $\mathit{sort}{=}\mathit{groan} \wedge \mathit{type}{=}\mathit{sound}$, will fail because it leads to an inconsistent knowledge base. One has both $\mathit{type}{=}\mathit{plant}$ and $\mathit{type}{=}\mathit{sound}$. The only way of blocking this result is to retract $\mathit{type}{=}\mathit{plant}$ from the knowledge base. However, this is not admissible because it violates the integrity constraint (or, from the point of view of an attribute-value structure, the appropriateness condition). We hypothesize the following relation between a parietal late positivity effect and our model of semantic processing.

(27) *Correlation parietal late positivity – belief revision:*
A parietal late positivity is triggered whenever an integrity constraint is violated s.t. a ‘normal’ revision operation is not applicable.

It seems that a sentence like ‘For the snowman’s eyes the kids used two pieces of coal. For his nose they used a groan from the fridge’ is interpretable only if at least part of the sentence is not taken in its literal sense. For example, the word 'groan' could be used in such a way that it refers to a nose-like object which emits a groan-like noise when squeezed. The general interpretative strategy behind this example can be described as follows. The information provided by the head noun is *not* taken as specifying the sort of the frame (e.g., it is a groan) but rather as giving a particular property of the object denoted by that frame, e.g. the value of a sound quality. The task of making sense of such sentences, then, consists in finding a frame s.t. (a) it satisfies the appropriateness condition, and (b) it has an attribute whose value can be of the sort denoted by the head noun.

Parietal late positivity also shows that during online semantic processing conclusions derived from more specific information do not always win out. Though it is true that semantic features got after semantic recognition overwrite contrary semantic features got prior to semantic recognition, features imposed on the argument can never be overwritten. Formally, this is expressed by having such a constraint be true at all worlds in a CO-model.

On the account developed above, integration / composition is always done w.r.t. a consistent set of features that are either imposed, predicted prior to recognition or got after semantic recognition. As a consequence, integration/composition are sensitive to differences between words of the same syntactic category denoting objects of the same type.

Summarizing, we have arrived at the following correlations. Late positivity effects are triggered whenever a prediction must be given up. Frontal late positivities are correlated with wrong guesses which do not violate a sortal type restriction on an argument of the verb. This is the case for within-context-violations and for between-context-violations. In this case integration, defined by the revision operation *, is possible. By contrast, for parietal late positivity effects, a violation of such a sortal type restriction occurs. In such a case normal integration is not possible.

4 Comparison to other approaches

In this section we will compare our model with other approaches and discuss some possible objections. First, we summarize the main theses underlying our model.

1. The N400 effect is related to lexical retrieval of items in LTM. In particular, it is directly related to the number of additional features (attribute-value pairs) that must be activated compared to the features which have already been activated during prediction. Hence, the N400 is not related to integration and/or composition.
2. The two late positivities are related to integration/composition: in order to arrive at a consistent new semantic representation (or knowledge base), predictions that are incompatible with the information got by bottom-up processing have to be given up. Formally, this amounts to revising the predictions with the (true) bottom-up information.

3. When taken together, (1) and (2) yield the following model of semantic processing in the brain: The first stage, indexed by the N400, is related to semantic retrieval whereas the second stage is related to integration/composition, which consists in a revision component made up by a retraction followed by an addition.

Thesis (1) is incompatible with a widely held view of N400 effects: the *integration view*. We will therefore begin by providing a critical assessment of this view in section 4.1. According to thesis (2), the P600 is a semantic effect, However, this is at odds with the widely held view according to which it reflects syntactic violations and syntactic repairing. Evidence for our interpretation will be discussed in section 4.2.

4.1 Integration view of the N400

On the *integration view*, the amplitude of the N400 is related to the effort of integrating a word in the current context, i.e. in the semantic representation built up so far. On this interpretation, N400 effects are (i) post lexical, i.e. they occur after a word has (semantically) been recognized and (ii) result from a combinatorial process. After a word has been accessed in LTM, the task consists in combining the semantic representation of the prior context with the semantic representation of that word. Thus, at any moment during semantic processing the set of semantic features is solely built up by words that have already been processed (or recognized) and not by features of (expected) upcoming words farther down the sentence.

As noted by [Pyl], a general problem of this account of the N400 is that the notion of 'semantic integration' is usually not sharply defined. As was already shown in the introduction, according to (most) formal semantic theories, composing a word after accessing it with the previous context does not depend on the way it is semantically related to this context in detail but merely on its general syntactic and semantic type. Furthermore, it is usually not explained why semantic expectedness and/or relatedness should affect semantic integration. Besides these theoretical weaknesses, this account also faces a number of severe empirical problems. First, the mismatch between the set of semantic features pre-activated by a prior context and a within-context-violation is greater in a strongly constraining than in a weakly constraining context so that it should be more difficult to integrate a WCV, like 'pine', in a strongly constraining context compared to a weakly constraining context. As an effect, the N400 amplitude in a strongly constraining context should exceed that in a weakly constraining context, contrary to the empirical findings. Second, the *integration view* predicts that for two words which are synonymous the difficulty in integrating them should be the same since they are necessarily semantically related to the prior context in exactly the same way. This prediction is falsified by example (5) in section 2.2. Since 'a' and 'an' have exactly the same meaning, they should elicit the same N400 effects. However, the amplitude of the N400 was larger for 'an' compared to 'a'.

4.2 The P600 as a measure of multimodal updating processes

According to our model, the late positivity (P600) is related to semantic integration.[14] Some predictions that have been made have to be given up because they are incompatible with the (semantic) information provided by the target word. This integration view of the P600 seems to be at odds with the popular *syntactic view* of this ERP-component. On this view, the P600 is interpreted as indexing the difficulty of revising or repairing a syntactic analysis when the target word makes the sentence based on this analysis ungrammatical (see [BFH12, 135] and [GPKP10] for an overview).

The syntactic view of the P600 has been challenged by a number of empirical results. First, [KHGH00] (see also [BFH12]) compared sentences with long distance wh-dependencies.

(28) a. Emily wonders who the performers in the concert imitate ...
b. Emily wonders whether the performers in the concert imitate ...

Only for (28a), but not for (28b), a P600 was found. Since (28a) is neither syntactically ill-formed nor does it contain a garden-path, this effect has to be explained in a different way. [KHGH00] suggest that in this case this effect reflects a process of syntactic integration: the verb 'imitate' has to be linked to the wh-pronoun 'who' whereas no such additional operation is needed in the case of (28b). Thus, the P600 is related to integration and not only to repairing. In addition, the linking that is required can be interpreted as being semantic in nature. A further example of a semantic P600 effect is given by so-called bridging phenomena, [Bur06,Sch13].

(29) Yesterday, a PhD. student was *shot/killed/found dead* downtown. The press reported that the pistol was probably from army stocks.

Both for 'killed' and 'found dead', [Bur06] found a P600 effect but not for 'shot'. According to [Bur06], this can be explained by assuming that in the former two cases establishing a coherent discourse relation (say, elaboration) requires more inferential work. Again, this is a purely semantic (or pragmatic) task which is related to integrating new information into the semantic representation of the previous context. In their discussion of the findings by [RGF11] and [Bur06], [BFH12, 136] conclude that 'what their materials have in common is that they require additional processing (as compared to the control condition) in order to arrive at a coherent mental representation of what the speaker or writer meant to communicate'. The authors hypothesize that all P600 effects can be described in terms of the construction, revision, or updating of a mental representation of what is being communicated, [BFH12, 137]. They argue that on this account of the P600 the observed effect reflects the efforts in reworking an initial mental representation and not the revision of a syntactic analysis. Thus for them, the P600 reflects integration difficulty. This difficulty is determined by how much the current mental representation needs to be adapted to incorporate the current

[14] This section owes much to the review article [BFH12].

input. They summarize their view of the P600 as follows [BFH12, 138]: 'The P600 component is the brain's natural electrophysiological reflection of updating a mental representation with new information.' Hence syntactic complexities and violations elicit a P600 effect because they reflect difficulties in building up a coherent mental representation at the syntactic level. Even more important than [BFH12]'s account of the P600 effect is the way they relate the N400 to this ERP-component. According to them, the N400 reflects the retrieval of the meaning of a word from LTM, [BFH12, 128].

5 Summary

In this paper we showed how a semantic theory can be extended to incorporate a 'predictive' and a 'revision' component in order to account for neurophysiological data on online semantic processing. The general idea is to use a decompositional frame semantics based on typed attribute-value structures together with a set of default rules. Such rules are in part pragmatic because their use is context dependent. Yet, the information inferred is always part of the semantic representation of a concept in LTM since the context only determines which part of the frame representing the concept is activated.

References

Ber10. J.J.A. Van Berkum. The brain is a prediction machine that cares about good and bad – any implications for neuropragmatics? *Italian Journal of Linguistics*, 22:181–208, 2010.

BFH12. Harm Brouwer, Hartmut Fitz, and John Hoeks. Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446:127 – 143, 2012.

BH11. Giosuè Baggio and Peter Hagoort. The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26(9):1338–1367, May 2011.

BHN$^+$09. Jos J.A. Van Berkum, Bregje Holleman, Mante Nieuwland, Marte Otten, and Jaap Murre. Right or wrong?: The brain's fast response to morally objectionable statements. *Psychological Science*, 20:1092–1099, 2009.

Bou94. Craig Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68(1):87 – 154, 1994.

Bur06. Petra Burkhardt. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159 – 168, 2006.

DQK14. Katherine A. DeLong, Laura Quante, and Marta Kutas. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61:150 – 162, 2014.

DUK05. Katherine A. DeLong, Thomas Urbach, and Marta Kutas. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8:1117 – 1121, 2005.

FK99. 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.

FWODK07. Kara D. Federmeier, Edward W Wlotko, Esmeralda De Ochoa-Dewald, and Marta Kutas. Multiple effects of sentential constraint on word processing. *Brain Research*, 1146:75–84, 2007.

GP91. Moisés Goldszmidt and Judea Pearl. System-Z+: A formalism for reasoning with variable-strength defaults. In Thomas L. Dean and Kathleen McKeown, editors, *Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, July 14-19, 1991, Volume 1.*, pages 399–404. AAAI Press / The MIT Press, 1991.

GP92. Moisés Goldszmidt and Judea Pearl. Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In Bernhard Nebel, Charles Rich, and William R. Swartout, editors, *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 661–672. Morgan Kaufmann, 1992.

GPKP10. Ana C. Gouvea, Colin Phillips, Nina Kazanina, and David Poeppel. The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2):149–188, 2010.

KF11. Marta Kutas and Kara Federmeier. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1):621 – 647, 2011.

KHGH00. Edith Kaan, Anthony Harris, Edward Gibson, and Phillip Holcomb. The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2):159–201, 2000.

LPP08. Ellen F. Lau, Colin Phillips, and David Poeppel. A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9:920–933, 2008.

Pea90. Judea Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In Rohit Parikh, editor, *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, March 1990*, pages 121–135. Morgan Kaufmann, 1990.

Pet07. W. Petersen. Representation of concepts as frames. In Jurgis Skilters, Fiorenza Toccafondi, and Gerhard Stemberger, editors, *Complex Cognition and Qualitative Science*, volume 2 of *The Baltic International Yearbook of Cognition, Logic and Communication*, pages 151–170. University of Latvia, 2007.

PO14. Wiebke Petersen and Tanja Osswald. Concept composition in frames: Focusing on genitive constructions. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald, and Wiebke Petersen, editors, *Frames and Concept Types*, volume 94 of *Studies in Linguistics and Philosophy*, pages 243–266. Springer International Publishing, 2014.

Pyl. Liina Pylkkänen. N400. Website. Online available `http://www.psych.nyu.edu/pylkkanen/Neural_Bases/07_slides/18_N400.pdf`;.

RGF11. Stefanie Regel, Thomas C. Gunter, and Angela D. Friederici. Isn't it ironic? an electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, 23(2):277 – 293, 2011.

Sch13. Petra B. Schumacher. Content and context in incremental processing: "the ham sandwich" revisited. *Philosophical Studies*, 168(1):151–165, 2013.