Modelling the retrieval-integration model with frames

Data and central issue. In this talk we present a formal account of the neurophysiological Retrieval-Integration model of online semantic processing, [Brouwer et al.]. According to this model, incremental, word-by-word language processing proceeds in retrieval-integration cycles where each cycle is modelled by a function $process$ which maps a word form w_1 and a context to an updated context. $Process$, in turn, is the composition of two functions $retrieve$ and $integrate$. The former maps a word form and the prior context to the disambiguated meaning of the word form whereas the latter takes this meaning and the context and maps it to an updated context. Both functions are indexed by particular ERP-components. For retrieval, this is the N400 whereas the P600 indexes $integrate$.

In our frame model retrieval and integration are related to different levels of a context. A context is a hierarchical (frame) structure consisting of three levels: (i) the sequence of discourse referents, reflecting the order in which they have been introduced into the text, (ii) an ordered sequence of events (and states) that is the restriction of the sequence in (i) to objects of sort 'event', and (iii) a sequence of situation models. Retrieval (N400 effect) is related to the current situation model whereas integration (P600 effect) is related to the current event. The N400 effect and the P600 effect are quantized by surprisal w.r.t. the current situation model and the current event, respectively.

(1)
a. He spread the warm bread with *butter/socks*.
b. 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 *snowman / jacket / towel* in the front yard.
c. [context: a tourist checking into an airplane with a huge suitcase and a woman behind the check-in counter deciding to charge the tourist extra for the weight of the suitcase] Next, the woman told the *suitcase* ....
d. The restaurant owner forgot which customer the waitress had *served*.
e. The restaurant owner forgot which waitress the customer had *served*.
f. The exterminator inquired which neighbor the landlord had *evicted*.
g. The neighbor inquired which exterminator the landlord had *evicted*.

In (1-a) an N400 effect is triggered after the (critical) word 'socks' compared to the control 'butter'. By contrast, neither in (1-b) nor in (1-c) an N400 effect is triggered for 'jacket' or 'suitcase'. If the context is left out (e.g. the first two sentences in (1-b)), an N400 effect occurs. Similarly, for the pair (1-d) and (1-e) no N400 effect is observed for the critical word 'serve'. Finally, for the pair (1-f) and (1-g), an N400 is triggered for the latter but not for the former on 'evict'. For the P600 one has: whenever an N400 is triggered, it is followed by a P600 effect (biphasic N400/P600 effect), e.g. (1-g) triggers an P600 effect. From the no N400 examples only (1-b), (1-c) and (1-e) trigger a P600 effect whereas (1-d) and (1-f) trigger neither a N400 nor a P600 effect.

Two existing models. One common analysis of the data in (1) are so-called multi-stream models of language processing. A semantic stream takes the context as a bag of words and tries to combine them to a meaningful proposition. E.g., both in (1-d) and (1-e) the waitress, the customer and the serving event can be combined into a meaningful proposition according to which the waitress served the customer. An N400 effect is triggered if this attempt fails. Distinct from this stream is an algorithmic, structure-driven stream that includes information about thematic relations and other structural devices like case marking. If the results of the two streams conflict, an attempt at (structural) revision is made which is reflected in a P600 effect. Other accounts of the N400 effect define the meaning of a word in terms of (structured) sets of features and relate this effect to prediction error, i.e. the difference between the semantic features the model expects to encounter and those features that are actually encountered, [Rabovsky/McRae]. The greater the accordance, the more attenuated is the N400 amplitude.

There are two principle problems for these approaches. First, there are contexts in which no N400 effect is triggered and a semantic incompatibility exists between the verb and at least one of the realized arguments as in (1-b) and (1-c). In these cases it is neither possible to construct a meaningful interpretation by combining the (meanings of) the words in any conceivable way nor is there a large overlap between predicted and encountered features. Second, the relation between the two effects cannot be linked to a mismatch between the results of a purely semantic and a more structure-biased parsing of the context. A prime counterexample are contexts with a biphasic N400/P600 effect in which the operations at both levels fail and, as a result, two effects are triggered but the models only predict an N400 effect because the streams are not in conflict.

Enriching the context with world knowledge. Common to the above mentioned models is the assumption that (i) predictions are a function of the current context and (ii) that the context is restricted to bottom-up information in form of the sequence of (the meaning of the) words that have already been encountered in the text or discourse. This has the effect that world knowledge is not accounted for. Consequently, predictions are restricted to elements of the meanings of discourse referents already introduced and these predictions are tested entirely in terms of semantic compatibility between discourse referents already introduced and discourse referents encountered next. Specifically, the unit targeted by predictions is the current event. This shows up in the constraint on predictions that it must be possible to combine the words in such a way that they yield a meaningful interpretation where by 'interpretation' is meant a model of the current event. Likewise, prediction of semantic features is restricted to sortal information that is related to the current event.

What the data, rather, show is that the criterion of whether a meaningful or probable interpretation can be constructed is related to the P600 effect and, therefore, to the operation of integration. A P600 effect is triggered in case of (i) semantic incompatibility and (ii) improbable thematic role assignment between the sort of an argument and the sort of an event. Hence, our model is thus based on the following assumptions: integration is related to the current event in a context (or information state). Whenever it is not possible to construct a probable, and a fortiori meaningful, event model, a P600 effect is triggered. Predictions, by contrast, are related to situation models, which go beyond the propositional content conveyed by the words in a context and which essentially involve world knowledge ([Zwaan et al., see [Ahrendt/Demberg] on learning such models from texts]). For example, in (1-b) a wintery scene and in (1-c) an airport scenario is described. Situation models basically are complex events that are ordered into an (action) sequence, a set of individuals participating in these actions and a participancy relation that links each individual to at least one event in the action sequence. Situation models do not contain information about thematic relations. Rather, relational information about objects in the context generalizes over particular thematic relations in terms of the participancy relation. One reason for this modelling is that expectations w.r.t. sorts are not directly related to particular thematic relations. For example, an individual who has the role of waitress in a restaurant scenario can be related to various actions with different thematic relations assigned to her. Hence, predicting a particular thematic role in addition to the sort would be error-prone.

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 (1-b) the jackets are expected because they are related to the children in a (yet to be introduced) state of wearing. By contrast, for the current event of building, jackets are unexpected and even lead to (unsolvable) problems when trying to integrate them into the model for this event. Therefore, no N400 effect but a P600 one is triggered. Similarly, in (1-d) and (1-e) having objects of sort 'waitress' and 'customer' in a restaurant scenario makes an action like serve likely. However, at the level of the current event the assignment of actor (theme) to the customer (waitress) leads to an improbable event model. By contrast, relating an event of sort 'evict' to two objects of sort 'landlord' and 'exterminator' is not expected in a housing scenario because an object of sort 'tenant' (as opposed to 'exterminator') is expected.

Incremental Dynamics with Frames. Frames are elements of a separate domain D_f of frames. The other domains are $D_o = D_i \cup D_e$ and D_w. The domain D_o is the union of the domain D_i of individuals (children etc.) and the domain D_e of events. D_e is sorted by $Sort$ and SM. Elements of $Sort$ are basic sorts like 'serve' whereas sorts belonging to SM are sorts for situation models. D_w is the domain of possible worlds. Frames are related to the other domains in the following way. Each frame is a partial description of a particular object. 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 (functional) relations. Each frame belongs to a particular world.On D_f an information ordering $\sqsubseteq$ is defined. One has $f \sqsubseteq f'$ if f and f' are descriptions of the same object in the same world and f' contains at least the information about the object f contains ($\theta(f) \subseteq \theta(f')$). At the discourse level we use Incremental Dynamics enriched with frames (for ID see [vanEijck]). Possibilities are pairs $\langle c, w\rangle$ consisting of a stack c and a world w. A particular stack position is a pair $\langle o, f\rangle$ consisting of an object and a frame that stores the information about o got in the discourse so far. An information state is a set of possibilities. Situation models are complex events. Their frames have an attribute ACTION_SEQUENCE whose value is the ordered sequence of actions (events) occurring in this scenario. A second attribute is PARTICIPANTS whose value is a set of individuals. Each element of this set is related to at least one action or one other participant. The field of this relation is the value of the attribute PARTICIPANCY_RELATION. Inside Incremental Dynamics with Frames the current situation model is stored in a particular stack position SM, similar to the reference time or the speaker. Similarly, the current event is stored in another stack position, called CE.

Modelling the RI-model. In order to define surprisal, we need probability measures over properties of events, either of a basic sort $\sigma \in Sort$ (integration, P600) or of a sort $\sigma \in SM$ (retrieval, N400). However, information growth related to update operations triggered by lexical items in our approach is done at the level of frames. Therefore, we define probabilities at the level of frames. The property Q of events associated with a set of frames $F = \{f_1, \dots, f_n\}$ is given by $Q_{f_1} \cap \dots \cap Q_{f_n}$, where for each $f \in F$ with $\theta(f) = \{R_1, \dots, R_m\}$ the property

Q_f is given by $dom(R_1) \cap \ldots \cap dom(R_n)$ with $dom(R) = \{f \mid \exists e.\exists o.R(f)(e)(o)\}$. Hence, we define for each $\sigma \in Sort \cup SM$ and $w \in D_w$, a probability measure Pr_{σ_w} on subsets of $D_{f_{\sigma_w}}$. The Pr_{σ_w} only give the probability of a set of frames in a particular world w. We need the probability of a property Q for a set of worlds V, in particular for the whole set D_w. In order to define this probability we first define a probability measure Pr_w on subsets of D_w. The probability of a property Q for a set of worlds V is now defined as in (2), [Gaerdenfors].

(2) $P_V(Q) := \Sigma_{w \in D_w} Pr_{\sigma_w}(Q) \cdot Pr_w(\{w\})/Pr_w(V)$ provided $Pr_w(V) \neq 0$.

In the next step we relate the probability measures to discourse update operations. Let $Pr_{\sigma_w}(\cdot)$ be a probability measure over frames related to events of sort σ in world w. If $f \sqsubseteq f'$ with $\theta(f') = \theta(f) \cup \{R\}$, then there is a frame f'' s.t. $\theta(f'') = \{R\}$. The operation of adding the frame f'' to the frame f is denoted by $\oplus$. If $f' = f \oplus f''$, the conditional probability $Pr_{\sigma_w}(Q_{f'} \mid Q_f)$ is the probability that frame f of sort σ can be extended to the frame f' of sort σ by adding frame f'' as a subframe to f. The relation to update operations is the following. If f is the frame component associated with the current situation model or the current event after processing the words $w_1 \ldots w_k$ and if f'' is a(!) frame associated with (the interpretation of) the word w_{k+1}, $Pr_{\sigma_w}(Q_{f'} \mid Q_f)$ is the probability relative to sort σ of encountering w_{k+1} given the context $w_1 \cdots w_k$ relative to a possibility p. Using a decompositional event semantics the interpretation of a word w possibly yields more than one decompositional predicate. As an effect, more than one frame is introduced which can be related to different elements of the context.

We follow [Champollion] and assume that the interpretation of a verb in the lexicon does not (yet) provide information about thematic relations. Rather, thematic roles are introduced separately. Specifically, we assume the following structure for DPs: $[[Det N]_{DP_1}[TR]]_{DP_2}$, [Bott/Sternefeld]. Whereas N provides sortal information, TR assigns a thematic role by which the object introduced by the interpretation of Det is related to the event introduced by the interpretation of the verb. Thus, N and TR introduce different kinds of information and, therefore, different frames f_N and f_{TR} that target the current situation model and the current event differently. This distinction is reflected in the update operations that interpret the three constituents. Similar to dynamic semantics, Det is interpreted as a domain expansion operation that (non-deterministically) adds a discourse object to each possibility in the current information state. This operation is related to level 1 and has no effects on the current situation model and the current event. The interpretation of N triggers two update operations. The first is the standard eliminative update operation from dynamic semantics. All possibilities in which the new object does not satisfy the sortal information provided by N are discarded. For the remaining possibilities, a second update operation is executed. The new object is added to the current situation model, i.e. to the value of the PARTICIPANCY attribute. As a result, its frame component is updated. Importantly, there is (yet) no change w.r.t. the current event because so far no thematic relation specifying how this object is related to this event has been determined. This is done by the update operation associated with TR. This operation always targets the current event (and not the current situation model) by updating its frame component with the attribute TR. The interpretation of verbs adds an event to the value of the ACTION_SEQUENCE attribute of the current situation model and (eliminatively) updates the current event's frame component with sortal information.

Surprisal is defined as follows. If $\theta(f) = \{R_1, \ldots, R_n\}$ for f the frame component of the current situation model or the current event in the input context, f is related to the frames $f_1, \ldots, f_n$. Here, each f_i is related to the (decompositional) interpretation of a word in the input context which contributes the property Q_{f_i}. Hence, the f_i form a sequence $f_1 f_2 \ldots f_n$. Finally, in order to define surprisal we have to consider not the Pr_{σ_w} but their lift to the level of D_w Pr_{D_w}. The surprisal of f'' is then defined by (3).

(3) $surprisal(Q_{f''}) = -log Pr_{D_w}(Q_{f''} \mid Q_{f_1} \cap \ldots \cap Q_{f_n})$.

Let us illustrate. For (1-d) and (1-e), both texts yield the same situation model after processing the initial segment preceding 'serve' because in both contexts the same sorts of objects have been introduced. So they do not differ relative to surprisal upon encountering 'serve' at this level and since we have already objects of sort 'waitress' and 'customer' a serving event gets a high probability relative to the information ordering $\sqsubseteq_{restaurant}$ and surprisal is thus low. By contrast, at the level of the current event the information is different because the assignments of sorts to thematic roles differ. As a result, the two corresponding frames occupy different positions relative to $\sqsubseteq_{Sort}$ involving a waitress and a customer. For the prototypical case, the waitress served the customer, the transition to the sort 'serve' is probable whereas for the reverse case this does not hold. Rather, actions like ask or complain are expected. Hence, a higher surprisal in the latter case. For (1-g), before encountering 'evict' the housing situation model contains participants of sort 'exterminator' and 'landlord'. A $\sqsubseteq_{housing}$ transition to an event of sort 'evict', as opposed to 'hire' is unlikely. At the level of the current event one has: given the two sorts with the assignment 'theme' and 'actor', respectively, a $\sqsubseteq_{Sort}$ transition to 'evict' is not probable. Hence, a high surprisal at both levels.

The contrast to (1-b) is that in the case of (1-b) the jackets can be related to the children (value of attribute CLOTHES) and jackets satisfy the constraint 'keep_warm' imposed by the context.

Summarizing we can say that the main aim of our presentation is to arrive at a formal semantic model that better complies with results from neurocognitive research. At the same time we also attempt to improve on neurocognitive theorizing by developing a formal model of the Retrieval_Integration model. Specifically, the functional roles of the N400 and the P600 effect are modelled by relating them to operations on the situation model and the current event, respectively. Retrieving the meaning of a word is modelled as the operation of introducing an object satisfying the meaning description in the current situation model. This operation is non-combinatorial because it does not relate this objects to the model of the current event. Rather, this operation triggers a transition in the frame hierarchy related to the current situation model. By contrast, the integration operation is always related to the model of the current event. This operation is therefore related to transitions in the frame hierarchy of the current event. From this functional characterization of the two effects the following prediction regarding the P600 effect can be derived. A necessary condition for this effect to be triggered is that the verb, denoting the current event, has already been processed.

Selected references: [Ahrendt/Demberg] Improving event prediction by representing script participants, *Proc. ACL* [Brouwer et al.] A neurocomputational model of the N400 and the P600 in language processing, *Cog.Sci.* • [Champollion] The interaction of compositional semantics and event semantics, *LP* • [Rabovsky/McRae] Simulating the N400 ERP-component as sem. network error, *Cognition*