

ESSLI



Twentieth



European Summer School



2008

**in Logic, Language
and Information**



Logic, language, and the brain

Michiel van Lambalgen

ESSLLI 2008

20th European Summer School in Logic, Language and Information

4–15 August 2008

Freie und Hansestadt Hamburg, Germany

Programme Committee. Enrico Franconi (Bolzano, Italy), Petra Hendriks (Groningen, The Netherlands), Michael Kaminski (Haifa, Israel), Benedikt Löwe (Amsterdam, The Netherlands & Hamburg, Germany) Massimo Poesio (Colchester, United Kingdom), Philippe Schlenker (Los Angeles CA, United States of America), Khalil Sima'an (Amsterdam, The Netherlands), Rineke Verbrugge (**Chair**, Groningen, The Netherlands).

Organizing Committee. Stefan Bold (Bonn, Germany), Hannah König (Hamburg, Germany), Benedikt Löwe (**chair**, Amsterdam, The Netherlands & Hamburg, Germany), Sanchit Saraf (Kanpur, India), Sara Uckelman (Amsterdam, The Netherlands), Hans van Ditmarsch (**chair**, Otago, New Zealand & Toulouse, France), Peter van Ormondt (Amsterdam, The Netherlands).

<http://www.illc.uva.nl/ESSLLI2008/>
esslli2008@science.uva.nl



INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

ESSLLI 2008 is organized by the Universität Hamburg under the auspices of the *Association for Logic, Language and Information* (FoLLI). The *Institute for Logic, Language and Computation* (ILLC) of the *Universiteit van Amsterdam* is providing important infrastructural support. Within the Universität Hamburg, ESSLLI 2008 is sponsored by the Departments *Informatik*, *Mathematik*, *Philosophie*, and *Sprache, Literatur, Medien I*, the *Fakultät für Mathematik, Informatik und Naturwissenschaften*, the *Zentrum für Sprachwissenschaft*, and the *Regionales Rechenzentrum*. ESSLLI 2008 is an event of the *Jahr der Mathematik 2008*. Further sponsors include the *Deutsche Forschungsgemeinschaft* (DFG), the Marie Curie Research Training Site GLoRiClass, the European Chapter of the Association for Computational Linguistics, the *Hamburgische Wissenschaftliche Stiftung*, the Kurt Gödel Society, Sun Microsystems, the Association for Symbolic Logic (ASL), and the European Association for Theoretical Computer Science (EATCS). The official airline of ESSLLI 2008 is Lufthansa; the book prize of the student session is sponsored by *Springer Verlag*.

Michiel van Lambalgen

Logic, language, and the brain

Course material. 20th European Summer School in Logic, Language and Information (ESSLLI 2008), Freie und Hansestadt Hamburg, Germany, 4–15 August 2008

The ESSLLI course material has been compiled by Michiel van Lambalgen. Unless otherwise mentioned, the copyright lies with the individual authors of the material. Michiel van Lambalgen declares that he has obtained all necessary permissions for the distribution of this material. ESSLLI 2008 and its organizers take no legal responsibility for the contents of this booklet.

LANGUAGE, LINGUISTICS AND COGNITION

CONTENTS

1. Two families of problems	2
2. Language and cognition	4
2.1. Cognitive substrates of tense and aspect	4
2.2. Planning, causality and the ordering of events	4
2.3. Why language matters to cognitive science	11
3. Linguistics and cognition	14
3.1. A paradox regained	14
3.2. The vagaries of intuition	17
3.3. Beyond competence and performance	22
3.4. Experiments on semantic processing	28
4. The road ahead	41
4.1. The binding problem for language	41
4.2. The semantic binding problem: toward a model of semantic unification	43
5. Conclusion	45
References	46

1. TWO FAMILIES OF PROBLEMS

Language has profound and intricate relations with other cognitive skills, enabling some while requiring others. Our ordinary experience might be sufficient to reveal a number of important aspects of the involvement of language in mental life. However, the variety of opinions and pieces of evidence discussed in the literature depicts a much more complex scenario than even an accomplished phenomenologist could discover. This chapter offers a partial overview of the debate on language and cognition over the past few decades. The landscape which we will try to render is vast and scattered, therefore it might be useful to start sketching the main lines using broad brush strokes. The premise of our discussion is a distinction between two families of problems, which we will refer to as ‘language and cognition’ and ‘linguistics and cognition’. We introduce these two clusters of issues below to discuss them in more detail in sections 2 and 3.

Under the first heading - language and cognition - we consider two problems. The first is which cognitive systems or functions support language. Hypotheses from linguistics, psychology and neuroscience take the view that language is anchored in such domains as perception and action. For instance, cognitive semanticists have suggested that meanings of concrete nouns stored in memory include stereotyped visual-geometric descriptions of objects [38]. Similarly, representations of action verbs might embrace some aspects of the relevant motor programs [26, 66, 67, 68]. It has been argued that even more basic ingredients of linguistic semantics like the predicate-argument structure derive from the functional and anatomical organization of the visual and auditory systems [37]. Language might as well exploit the resources of higher cognitive domains. Recent proposals suggest that planning supports to some extent the representations and computations involved in syntax and semantics. The recursive structure of plans supplies a basis for combinatory operations in grammar [78], while the goal-directed nature of planned actions constrains cognitive constructions of time, causality and events, with important consequences for the semantics of tense and aspect [78, 90]. Discourse processing, requiring the recomputation of meanings when unexpected information is provided, might operate on the basis of the defeasible logic underlying planning, allowing for action sequences to be readjusted when obstacles are encountered on the way [81, 90]. We shall return to the issues of cognitive time, tense and aspect and the logic of discourse processing in several occasions in this chapter.

The second problem, which has captured the interests of many philosophers, is which cognitive functions are triggered or facilitated by language. Language has been argued to boost a number of social and cognitive phenomena, the most striking of which is perhaps the speed of cultural transmission during both phylogeny and ontogeny [16]. Particularly relevant for the purposes of this chapter is the more specific issue of the dependencies between interpretation and reasoning. Some have seen non-domain-specific thought and reasoning [7, 8] as the most prominent of the mental skills subserved by language. This idea is explicit in the tradition of logical approaches to language at least since Boole¹ and bears some resemblance to the psycholinguistic notion that interpretation (computing truth-values) has logical and temporal priority over reasoning (computing with truth-values) [41]. On this view, inference not only elaborates or enriches propositional meanings [76], but also establishes a number of conceptual links with other representations, such as world knowledge, action schemes, and so on. Linguistic interpretation thus supports reasoning,

¹In his *Laws of Thought* Boole writes: “That language is an instrument of human reason, and not merely a medium for the expression of thought, is a truth generally admitted. It is proposed in this chapter to inquire what is that renders language thus subservient to the most important of our intellectual faculties.” [3, p. 24]

but not vice-versa. Others have seen the relation between inference and interpretation as reversed, if not based on an interaction [81, 82]. More precisely, human communication is regarded as the most sophisticated skill enabled by language while reasoning serves the purpose of coordinating different interpretations of an utterance or different situation models [80]. We shall touch upon the issue of reasoning and interpretation in sections 2.3, 3.4 and 4.2. In section 2.3 we will try to exemplify the claim that language matters to cognitive science discussing the consequences of linguistic interpretation on reasoning tasks.

Under the second heading - linguistics and cognition - we include two broad questions. The first is whether the structures and rules described in linguistics have any psychological reality. There exist several opinions in this regard, occupying different positions on the spectrum between mentalism and anti-mentalism. At one extreme lies cognitive linguistics [15, 34, 90], endorsing both methodological and theoretical mentalism. As far as the latter is concerned, cognitive linguistics sees phonological, syntactic and semantic structures as representations entertaining causal and conceptual relationships with other mental entities. Proponents of cognitive approaches to language have also called for a revision of linguistic methodology, favouring a higher degree of interaction with psychology and neuroscience [18]. At the opposite side of the spectrum lies formal semantic theory which, inspired by Frege's anti-psychologistic stance on meaning and thought [5], rejects theoretical as well as methodological mentalism. Researchers in the tradition of Frege and Montague typically view linguistics as separate from the study of the social and cognitive aspects of language [51]. As has been recently argued [34], however, there is no opposition between the tenets of cognitive linguistics and the use of model-theoretic or other formal tools to describe semantic structures and computations. As for theoretical mentalism, formal semanticists view meanings as objective entities, distinguished from their subjective representations.² Between the two poles is Chomsky's [11] theoretical mentalism viewing syntactic principles as ultimately residing in the mind-brain of language users. Still, in generative grammar a commitment to the psychological reality of language does not lead to criticism of standard linguistic methodology, which is maintained in its traditional form based on intuitions and the competence-performance distinction. In sections 3.1 and 3.2 we shall argue that these are among the heaviest burdens for a cognitive-computational approach to language.

The second question, not independent of the first, is whether experimental data from language processing have any bearing on linguistic theory. On this issue, the answers of linguists have been mostly negative. Chomsky's competence-performance distinction, in particular, has been used to secure linguistics from experimental evidence of any sort, while intuitions are the only type of empirical data supported by the theory. More recently, however, some commentators have argued that behavioral data should also be allowed to affect theories of competence, especially when the linguist is studying a language which is not her own native language [54]. Others, for instance Jackendoff [39], have proposed novel theoretical frameworks in which accounts of competence are connected to possible

²In *On sense and reference* Frege writes: "The reference and sense of a sign are to be distinguished from the associated idea. If the reference of a sign is an object perceivable by the senses, my idea of it is an internal image, arising from memories of sense impressions which I have had and acts, both internal and external, which I have performed. Such an idea is often saturated with feeling; the clarity of its separate parts varies and oscillates. The same sense is not always connected, even in the same man, with the same idea. The idea is subjective: one man's idea is not that of another. There result, as a matter of course, a variety of differences in the ideas associated with the same sense. A painter, a horseman, and a zoologist will probably connect different ideas with the name 'Bucephalus'. This constitutes an essential distinction between the idea and the sign's sense, which may be the common property of many and therefore is not a part of a mode of the individual mind." [21, pp. 24-25]

performance mechanisms. Although these models seem to account for how competence constrains processing, they are blind to the possibility that competence is the result of an optimal adaptation for performance. In this sense, most of current proposals within both generative grammar and formal semantics view the theory of competence as shielded from experimental data of the kind that psycholinguistics and the cognitive neuroscience of language customarily provide. The issue of competence theories and experimental data is of paramount importance for our purposes and shall be discussed in sections 3.1, 3.2 and 3.3. A more liberal attitude has been adopted on the other side of the fence by psycholinguists and neuroscientists. In several studies, specific linguistic hypotheses have been discussed in the light of experimental data, if not explicitly tested [6, 19, 59, 60]. Despite these fine attempts, which undeniably constitute a progress, unified frameworks accommodating linguistic theories, computational accounts of parsing or interpretation and psychological processing models based on behavioral, electrophysiological and neuroimaging data are still missing. The aim of this chapter is to remove a few obstacles on the road and take some tentative steps in that direction.

2. LANGUAGE AND COGNITION

2.1. Cognitive substrates of tense and aspect. We take it as the essential purpose of tense and aspect to facilitate the computation of the structure of the events described in a narrative. One implication of this characterization is that, contrary to what generative approaches to linguistics maintain it is not very useful to study tense and aspect at the sentence level only; tense and aspect really come into their own only at the discourse level. However, tense and aspect cannot by themselves completely determine event structure, and must recruit world knowledge. The following examples (2.1) will make clear what we have in mind. French has several past tenses (Passé Simple, Imparfait, Passé Composé), which differ in their aspectual contribution. The following mini-discourses in French all consist of one sentence in the Imparfait and one in the Passé Simple. The structure of the set of events differs in each case, however.

- (1) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)
It was hot. Jean took off his sweater.
 b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)
Jean got a ticket. He was driving too fast.
 c. Jean appuya sur l'interrupteur. La lumière l'éblouissait. (PS, Imp)
Jean pushed the button. The light blinded him.

In the first case, the Imp-sentence describes the background against which the event described by the PS-sentence occurs. In the second case, the event described by the PS terminates the event described by the Imp, whereas in the third case the relation is rather one of initiation. These examples also show that world-knowledge in the form of knowledge of causal relationships is an essential ingredient in determining event structure. This knowledge is mostly applied automatically in computing event structure, but may be consciously recruited if the automatic processing leaves the event structure still underdetermined. It is the task of cognitive science to determine what this algorithm is, and how it is actually implemented. The important role of causal relationships highlighted in the above examples suggests a direction in which to look for that algorithm.

2.2. Planning, causality and the ordering of events. Stated bluntly, our main hypothesis is:

the ability to automatically derive the discourse model determined by a narrative is subserved by the ability to compute plans to achieve a given goal.

We present several converging lines of evidence which lend some plausibility to this conjecture. First, a distinguishing feature of human cognition is that it is goal-oriented, with goals ranging from very short-term (get a glass of water) to very long-term (having sufficient income after retirement). In each case, the goal comes with a plan which produces an ordered collection of actions, be they motor actions or transfers of money to a special account. The two cases differ in that the former plan is for the most part generated automatically, whereas the latter plan may involve a good deal of deliberation. More precisely, planning consists in

the construction of a *sequence*³ of actions which will achieve a given goal, taking into account properties of the world and the agent, and also events that might occur in the world.

In the literature there have been various attempts to link the language capacity with the planning capacity. The setting is usually a discussion of the evolutionary origin of language. Even if it is granted that some nonhuman primates have learned a primitive form of language, there is still a striking difference in language proficiency between chimpanzees and ourselves. It is still a matter of ongoing debate to determine exactly what the difference consists in. Some would say that the difference is in syntax: human syntax is recursive, the chimpanzee's syntax (if that is the word) is not. One may then point to an analogy between language and planning. Language production can be characterized as transforming a semantic structure, to which the notion of linearity may not be applicable, into linear form (the linguistic utterance); as we have seen, planning also involves linearization. In the next step of the argument, the recursive structure of syntax is then linked to the recursive structure, i.e. the hierarchical organization, of planning (see for example Greenfield [25] and Steedman [79]). That is, planning is used to explain both the continuity with nonhuman primates, and the divergence. Both humans and nonhuman primates engage in planning. Primates are adept at planning (at least for time-spans not exceeding 48 hours), as has been known since Kohler's 1925 observations [44]. It has even been attested in monkeys, in recent experiments with squirrel monkeys by McGonigle, Chalmers and Dickinson [58].

Finally, a more direct link between temporal language and planning was investigated experimentally by Trabasso and Stein [84] in a paper whose title sums up the program: 'Using goal-plan knowledge to merge the past with the present and the future in narrating events on line'. The paper defends the thesis that 'the plan unites the past (a desired state) with the present (an attempt) and the future (the attainment of that state) [84, p. 322] and "[c]ausality and planning provide the medium through which the past is glued to the present and future" [84, p. 347]. Trabasso and Stein present the results of an experiment in which children and adults were asked to narrate a sequence of 24 scenes in a picture storybook called *Frog, where are you?*, in which a boy attempts to find his pet frog which has escaped from its jar⁴. The drawings depict various failed attempts, until the boy finds his frog by accident. The purpose of the experiment is to investigate what linguistic devices, in particular temporal expressions, children use to narrate the story, as a function of age. They provide some protocols which show a child of age 3 narrating the story in

³More complex plans are possible which also involve overlapping actions, such as for example drinking while walking.

⁴This is a classic experimental paradigm for investigating the acquisition of temporal notions in children. See Berman and Slobin [1] for methods, results, and last but not least, the frog pictures themselves.

tenseless fashion, describing a sequence of objects and actions without relating them to other objects or actions; none of the encoded actions is relevant to the boy's ultimate goal. Temporal sequencing comes at age 4, and now some of the encoded actions are relevant to the goal. Explicit awareness that a particular action is instrumental toward a goal shows up at age 5. At age 9, action–goal relationships are marked increasingly, and (normal) adults structure the narrative completely as a series of failed or successful attempts to reach the goal. We can see from this that there is a connection between children's developing sense of time, and their ability to structure the narrative as the execution of a plan toward the goal of finding the frog. The child of age 3 is glued to the present. The child of 4 includes causal relations between events, states of mind and actions; these causal relations implicitly drive the narrative forward. The child of 5 can move from narrating a current action to mentioning a goal state to be attained in the future, and back again. The reason that there must be a *gradual* development of these capabilities is outlined in the following quote

Inferring goals and plans involves considerable social and personal knowledge and places heavy demands on a narrator's working memory. The child who narrates events needs to attend to and maintain the current event in working memory; to activate and retrieve prior knowledge relevant to events, either in general or from earlier parts of the story, in order to interpret and explain the current event; and to integrate these interpretations into a context within a plan, all within the limitations of knowledge and working memory. In effect, over time the child is engaged in dynamic thinking, actively constructing and evaluating models and hypotheses about what is occurring. In so doing, the child creates a changing mental model that results in a long-term memory representation of what has occurred. [84, p. 327]

Working memory is thus essentially involved in this process of discourse integration, and failures in its operation may show up as deficiencies in the use of temporal language.

What is interesting for our purposes is that the ingredients that jointly enable planning each have a role to play in the construction of discourse models. Take causality, for instance, shown to be involved in the construction of discourse models corresponding to (2.1). Planning essentially requires knowledge of the causal effects of actions, as well as of the causal effects of possible events in the world; thus the planning capacity must have devised ways of retrieving such knowledge from declarative memory. Planning also essentially involves ordering actions with respect to each other and to events occurring in the world (i.e. not dependent on the agent); furthermore the resulting structure must be held in memory at least until the goal envisaged is reached. The reader can easily envisage this by the considering planning the steps which lead to a pile of pancakes – e.g. causal knowledge dictates that one must pour oil in the frying-pan before putting in the batter, and this knowledge has to remain active as long as one is not finished.

The fundamental *logical* connection between discourse interpretation and planning is that both are defeasible. When we plan (deliberately or automatically) we plan in virtue of our best guess about the world in which we will have to execute our plan. We may plan for what to do if we miss the bus, but we don't plan for what to do if the bus doesn't come because the gravitational constant changes, even though that is a logical possibility. Similarly, the computation of a discourse structure is defeasible: the reader who sees

(2) Bill used to be a member of a subversive organization.

is likely to infer (i.e. to read off from the discourse structure) that Bill is no longer a member, but that implicature can easily be cancelled, as in

(3) Bill used to be a member of a subversive organization, and he still is.

In cognitive terms, planning is part of ‘executive function’, an umbrella term for processes responsible for higher-level action control that are necessary for maintaining a goal and achieving it in possibly adverse circumstances; executive function comprises maintaining a goal, planning, inhibition, coordination and control of action sequences. Our main hypothesis can thus be formulated as: several components of executive function involved in comprehension and production of tense. A corollary of the hypothesis is that failures of executive function can show up in deviant use of verb tenses (e.g. in ADHD, or schizophrenia).

The link between planning and verb tenses is provided by the notion of *goal*: in both comprehension and production, the goal is to introduce the event corresponding to the tensed VP into the event structure. This goal must have two components:

- location of event in time
- meshing it with other events

An example will make this clearer: consider what goes on in comprehending the mini-discourse

(4) Max fell. John pushed him.

On one prominent reading, the event described by the second sentence precedes, indeed causes, that described in the first sentence. The relevant goals are in this case:

- ‘update discourse with past event $e_1 = fall(m)$ and fit e_1 in context’
- ‘update discourse with past event $e_2 = push(j, m)$ and fit e_2 in context’

Planning must determine the order of e_1, e_2 , and to do so, the planning system recruits causal knowledge, as well as the principle that causes precede effects.

To give the reader a clearer picture of what goes on in such computations, we have to introduce some notation⁵ which will also be useful for our discussion of neuroimaging of language and of the binding problem later in the chapter. We make a distinction between *events* (denoted e, e_1, e_2, \dots) and *processes* (denoted f, f_1, f_2, \dots). We say that events *occur* or *happen* at a particular time, and represent this formally by the expression *Happens*(e, t). By contrast, processes do not occur, but are *going on* at a particular time, and for this we use the predicate *HoldsAt*(f, t). Events and processes can stand in causal relations. For instance, an event may kick off a process (*Initiates*(e, f, t), or it may end it (*Terminate*(e, f, t)). We will use these predicates to mean the causal relation only; it is not implied that e actually occurs. Finally, a useful defined predicate is *Clipped*(s, f, t), which says that between times s and t an event occurs which ends the process f . The predicates just introduced are related by axioms, of which we will see a glimpse below.

With this notation, and using the further notation ‘... succeeds’ to abbreviate: ‘make it the case in our discourse model that ...’⁶, we can write the two update instructions involved in comprehending the discourse as

- $?Happens(e_1, t), t < now, Happens(e', t')$ succeeds
- $?Happens(e_2, s), s < now, Happens(e'', t'')$ succeeds

Here, e' and e'' are *variables* for event types in the context – which have to be found by substitution (more precisely, *unification*). These two update instructions have to be executed, preferably so that $e'' = e_1$ and $s < t''$ – if ‘Max fell’ is the first sentence of

⁵Borrowed from the Event Calculus [90].

⁶This notation derives from logic programming. By itself, $?\phi(x)$ denotes a *goal* or *query*, a request for a value a of x such that $\phi(a)$ is true. The answer may be negative, if the database against which $\phi(x)$ is checked contains no such individual. by $?\phi(x)$ succeeds we mean that in such cases the database must be updated with an a making ϕ true. These requests for updates are also known as *integrity constraints*.

the discourse, we may disregard e' .⁷ To formulate the causal knowledge relevant to the execution of these instructions we introduce a process f (*falling*) corresponding to the event e_1 ($fall(m)$), where f , e_1 and e_2 are related by the following statements

- $HoldsAt(f, t) \rightarrow Happens(e_1, t)$
- $Initiates(e_2, f, s)$

The hearer processing the discourse will first satisfy the update request corresponding to ‘Max fell.’ by locating the event e_1 sometime in the past. The second sentence, ‘John pushed him.’, is represented by the request

$$?Happens(e_2, s), s < now, Happens(e'', t'')$$

which contains the variable e'' . The hearer will try to satisfy the goal by reducing it using relevant causal knowledge. Thus, by applying 2.2 and putting⁸ $e'' = e_1 = fall(m)$, the second request is reduced to

$$?Happens(e_2, s), s < now, Happens(e_1, t''), HoldsAt(f, t'') \text{ succeeds.}$$

The system now applies a general causal principle about the world, known as ‘inertia’, which says that if an event e has kicked off a process f at time t , and nothing happens to terminate the process between t and t' , then f is still going on at t' . This principle rules out spontaneous changes, that is, changes which are not caused by the occurrences of events. This principle can be formulated as the axiom

$$Happens(e, t) \wedge Initiates(e, f, t) \wedge t < t' \wedge \neg Clipped(t, f, t') \rightarrow HoldsAt(f, t')$$

Using this axiom, the request is further reduced to

$$?Happens(e_2, s), s < now, Happens(e_1, t''), Happens(e, t), Initiates(e, f, t), \\ t < t'', \neg Clipped(t, f, t'') \text{ succeeds}$$

using 2.2 and putting $e = e_2 = push(j, m)$, $s = t$, we reduce this request to

$$?Happens(e_2, s), s < now, Happens(e_1, t''), s < t'', \neg Clipped(s, f, t'') \text{ succeeds}$$

This is a definite update request which almost says that *push* precedes *fall* – except for the formula $\neg Clipped(s, f, t'')$ which expresses that f has not been terminated between s and t'' . If f were terminated between s and t'' we have a situation as in

- (5) Max fell. John pushed him a second time and Max fell all the way to the bottom of the pit.

Since we have no information to this effect, we may assume $\neg Clipped(s, f, t'')$. This form of argument is known as *closed world reasoning*: ‘assume all those propositions to be false which you have no reason to assume to be true’. Closed world reasoning is essential to planning, because it allows one to discount events which are logically possible but in practice irrelevant. The final update request is thus

$$?Happens(e_2, s), s < now, Happens(e_1, t''), s < t'' \text{ succeeds}$$

which is the instruction to update the discourse model with the past events e_1 and e_2 such that e_2 precedes e_1 .

Now just as plans may have to be revised in mid-execution (it turns out there is not sufficient oil to produce the projected number of pancakes), discourse models may have to be re-computed. As will be seen below, when discussing temporal connectives, such re-computations might leave definite electro-physiological traces. We will elaborate our

⁷Thus we regard context here as provided by preceding discourse, although one may conceive of ‘forward-looking’ notions of context as well.

⁸This *unification* will be important in our discussion of the binding problem.

running example to give the reader an idea of what is involved. Suppose the discourse does not stop after ‘John pushed him.’, but instead continues:

(6) Max fell. John pushed him, or rather what was left of him, over the edge.

One prominent interpretation is now that $e_2 = \text{push}(j,m)$ comes after $e_1 = \text{fall}(m)$ – and this is also the result of a computation, in fact a re-computation, since after the first ‘him’ the hearer may have computed that e_2 precedes e_1 . Here is an informal sketch of this re-computation. The phrase ‘or rather what was left of him’ suggests Max is now *dead*, so the update request corresponding to the second sentence is something like

$?Happens(e_2, s), s < \text{now}, HoldsAt(\text{dead}(m), s), Happens(e'', t'')$ succeeds,

perhaps together with a requirement to the effect that the entire *pushing* event occurs while *dead(m)* obtains. It now seems reasonable to assume that at the start of *falling* (the process f), Max is still *alive*. Again unifying e'' and e_1 and applying property 2.2 the request reduces to finding instants s, t'' such that

$?Happens(e_2, s), s < \text{now}, HoldsAt(\text{dead}(m), s), HoldsAt(\text{alive}(m), t''), Happens(e_1, t'')$ succeeds

can be satisfied. Since *alive* always precedes *dead* and not conversely, it follows that we must have that $e_1 = \text{fall}$ precedes $e_2 = \text{push}$.

In summary, what is suggested here is a universal computational mechanism for determining event structures, based on the planning system. Temporal expressions (not only tense, but also temporal prepositions as will be seen below) are hypothesized to determine requests for an update of the current discourse model. Processing these requests involves unification and search through semantic memory, as well as setting up temporary structures in working memory. These computational processes are likely to leave traces on the EEG, a subject we will explore in greater depth below.

2.2.1. *Computing event structures for (PS, Imp) combinations.* The same type of argument works for the French examples with which we started this section.

(1) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)

It was hot. Jean took off his sweater.

Intuitively, this narrative determines an event structure in which ‘hot’ acts as a background which is true all the time; the foregrounded event (‘taking off one’s sweater’) is placed inside this background. One arrives at this structure by means of the following argument. World knowledge contains no causal link to the effect that taking off one’s sweater changes the temperature. The goal corresponding to the first sentence dictates that it is hot at some t before *now*. By the principle of inertia, the state *hot* must either hold initially (i.e. at the beginning of the period described by the discourse) or have been initiated. The latter requires the occurrence of an event, which is however not given by the discourse. Therefore *hot* holds initially. Similarly no terminating event is mentioned, so that *hot* extends indefinitely, and it follows that the event described by the second sentence must be positioned inside *hot*.

The second example

(1) b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)

Jean got a ticket. He was driving too fast.

dates from the bygone days when speeding cars were stopped by the police instead of being photographed. It is given that the event of getting a ticket occurred sometime in the past. It is also given that the fluent *speeding* was true some time in the past, hence it holds initially

or has been initiated. We have to determine the relative position of event and fluent. World knowledge yields that getting a *ticket* terminates, but not initiates, *speeding*. Since this is the only event mentioned, *speeding* holds from the beginning of discourse, and is not re-initiated once it has been terminated.

In the final example (1c), the same order of the tenses yields a different event order, guided by the application of causal knowledge.

- (1) c. Jean appuya sur l'interrupteur. La lumière l'éblouissait. (PS, Imp).
Jean pushed the button. The light blinded him.

One (occurrence of an) action is mentioned, *pushing* the light button, which has the causal effect of initiating the light being on when its current state is off. No terminating event is mentioned, so that the light remains on. It also follows that the light must be off for some time prior to being switched on, and therefore that it must be off at the beginning of discourse. The definite article in '*La lumière*' leads to a search for an antecedently introduced light, which successfully terminates after unification with the light introduced in the first sentence; therefore it is this light which is too bright.

2.2.2. *Deviant verb tenses and ADHD*. The preceding considerations can be applied to psychiatric disorders which are known to involve language impairments, sometimes of a very subtle kind. For instance, children with ADHD have difficulties with retelling a story, a task which involves presenting information so that it is organized, (temporally) coherent and adjusted to the needs of the listener. The ability to attend to these requirements while performing the task presupposes retaining goals in working memory, while planning the necessary steps and monitoring their executive – in short, this ability requires executive function as defined above [69].

The present model can be used to derive predictions on deviant narration in children with ADHD. It is known that these children are not very good at keeping goals active in working memory [22]. Now recall that the update request, i.e. the goal to be satisfied, corresponding to a verb tense consists of two components

- location of an event in time
- meshing the event with other events

If a child has trouble maintaining a goal in working memory, this may lead to a simplification of that goal. In the case of verb tenses the most likely simplification is to 'location of event in time' (never mind the meshing with other events), since this involves the least processing (search through semantic memory and unification). This simplification affects both comprehension and production, the case of interest here. Indeed, in producing speech which is attuned to the needs of the listener, the speaker must construct a discourse model of his own utterances, to determine whether it is sufficiently unambiguous. Now a typical finding in this area is

ADHD groups exhibited a higher frequency of sequence errors, which reflects a breakdown in the global organization of story theme. The pure-ADHD group [i.e. the group without reading disability] had difficulties in local organization, reflected in ambiguous references [of pronouns referring to persons or events]. These are failures in cohesion which make it difficult for the listener to follow the speaker's train of thought. Ambiguous references can result from a failure to organize and monitor the cohesion of sentences, as well as from a failure to take into account the needs of the listener. [69, p. 141]

This seems to suggest that there is indeed an intimate connection between language comprehension and production, and processing of goals.

2.3. Why language matters to cognitive science. All experiments on human subjects in cognitive science require verbal instructions, but those in which stimulus and/or response are verbal are of special interest to us, because here the interpretation of the experiment completely depends on the experimenter's interpretation of what the subject does with the verbal material. As a number of examples will show, experimenters have often been too complacent in assuming that their interpretation coincides with the subjects'. Our examples are drawn from the psychology of reasoning. Results in the psychology of reasoning have often been viewed as implying that logical reasoning, if not exactly a party trick, is an ability acquired only in the course of becoming literate, for instance by going to school, or worse, by attending a logic class; and that even then deviations from the norms of classical logic abound.⁹ We believe that these views betray a lack of awareness of the tremendous variation in language understanding and especially language use. The following sections provide a number of examples.

2.3.1. Reasoning in illiterate subjects. A fundamental observation about explained about reasoning in illiterate subjects is their obstinate refusal to draw conclusions from premisses supplied by the experimenter, as in Luria's famous example [53]

E. In the Far North, where there is snow, all bears are white. Novaya Zemlya is in the Far North and there is always snow there. What colour are the bears there?

S. I don't know what colour the bears are there, I never saw them.

... *E.* But what do you think?

S. Once I saw a bear in a museum, that's all.

E. But on the basis of what I said, what colour do you think the bears are there?

S. Either one-coloured or two-coloured ... [ponders for a long time]. To judge from the place, they should be white. You say that there is a lot of snow there, but we have never been there!

Here Luria is talking to an illiterate peasant from Kazakhstan. He attributed the peasant's refusal to answer, to an overriding need to answer from personal knowledge only, interfering with the deductive inference.

A careful analysis of these and related data shows, however, that the situation is considerably more complicated, and that Luria failed to take into account the pragmatics of language use. For one thing, the subject is not simply refusing to answer. He *does* draw the inference when he says "To judge from the place, they should be white"; but he refuses to consider this an answer to the question posed by the experimenter. This has to do with pragmatic aspects of questions and answers in natural language. Why should someone be interested in truth *relative* to assumptions, instead of absolute truth? The subject may refuse to answer only because he does not know whether the premisses are really true; not because of an inability to draw the correct inference. This point can be amplified in several ways.

Sometimes the refusal to answer is motivated by a piece of logical (meta-)reasoning, as in Scribner's and Cole's study of reasoning among the illiterate Kpelle tribe in Liberia (see [75]). Here is a sample argument given to her subjects

All Kpelle men are rice farmers.

Mr. Smith¹⁰ is not a rice farmer.

Is Mr. Smith a Kpelle man?

Consider the following dialogue with a subject:

⁹See [83] for discussion.

¹⁰'Mr. Smith' is not a possible Kpelle name.

- S. I don't know the man in person. I have not laid eyes on the man himself.
 E. Just think about the statement.
 S. If I know a man in person, I can answer that question, but since I do not know him in person I cannot answer that question.
 E. Try and answer from your Kpelle sense.
 S. If you know a person, if a question comes up about him you are able to answer. But if you do not know a person, if a question comes up about him, it's hard for you to answer it.

Under the plausible assumption that in the dialogue 'if' means 'if and only if', the Kpelle subject actually makes the *modus tollens* inference in the meta-argument (in his second turn) that he refuses to make in the object-argument! Luria had concluded that subjects' thinking is limited to the concrete and situational, that of which they have personal knowledge: Scribner tended to the conclusion that subjects fail because they don't understand the problem, or what kind of discourse 'game' is intended. If this means that subjects do not understand what the experimenter wants from them, one can only agree, as long as this is not taken to imply that there is some deficiency in the subjects' logical reasoning. Many of these examples bear the hallmarks of moral scruples about bearing witness on the basis of hearsay.

We may note here that the experimenter intends the subject (a) to understand the task ('assume only these two premisses, nothing more'), (b) to interpret the conditional as the material implication, and (c) to keep in mind the literal wording of the premisses. We will discuss (b) in section 2.3.2. As regards (a), task understanding, it may very well be that this requires understanding of a particular discourse genre, roughly that of an exam question in physics, which is not available to the unschooled. Imagine, for example, the confusion the subject gets into if he interprets the experimenter's question as a sincere request for information! It may also be difficult (at least for illiterates) to memorise the literal wording of the premisses, instead of the mental image they convey. This of course makes inference from 'given premisses' a non-starter. Here is another example from Luria [53].

- E. Cotton can only grow where it is hot and dry. In England it is cold and damp. Can cotton grown there?
 S. I don't know
 :
 S. I've only been in Kashgar country; I don't know beyond that.
 E. But on the basis of what I said to you, can cotton grown there?
 S. If the land is good, cotton will grown there, but if it is damp and poor, it won't grow. If it's like the Kashgar country, it will grow there too. If the soil is loose, it can grow there too, of course.

Luria interpreted this phenomenon as a refitting of the premisses according to convention; it seems better to describe it as the construction of a discourse model of the premisses using general knowledge ('cold and damp implies poor soil') and solving the task by describing the model, in line with the famous experiments on sentence memory by [4].

2.3.2. *Rational reasoning and the material implication.* No experiment in the psychology of reasoning has generated more controversy, or more theories of reasoning, than the Watson Selection Task, of which figure 1 is an example. The reader, who has probably seen the task before, should realise that this is all the information provided to the subjects, in order to appreciate the tremendous difficulty posed by this task. This experiment has been replicated many times. If one formulates the rule

Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which if any of these four cards you *must* turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the cards you want to turn.

Rule: *If there is a vowel on one side, then there is an even number on the other side.*

Cards:



FIGURE 1. Wason's selection task

p	p, q	$p, \neg q$	$p, q, \neg q$	misc.
35%	45%	5%	7%	8%

TABLE 1. Typical scores in the selection task

If there is a vowel on one side, then there is an even number on the other side.

as an implication $p \rightarrow q$, then the observed pattern of results is typically as given in table 1. Thus, the modal response (around half of the undergraduate subjects) is to turn A and 4. Very few subjects turn A and 7. Wason (and, until fairly recently, the great majority of researchers) assumed, without considering alternatives, that the correct performance is to turn the A and 7 cards only. This led him to attribute 'irrationality' to his subjects, and others to deny that formal logic is relevant to actual human reasoning (see [83] for discussion and references).

In a very real sense, however, Wason got his own task wrong in stipulating that there was a particular 'obviously correct' answer. The cards corresponding to p and $\neg q$ form the correct choice on the assumption that the expression 'if ... then' in the rule is interpreted using the material implication. However, nothing in the experimental material itself forces this interpretation, and the experimenter must therefore assume that the subject comes to the experiment equipped with this interpretation. Readers of this handbook need not be reminded that conditionals in natural language do not have a uniform interpretation, and can rarely be represented as the material implication. For example, conditionals are often interpreted as exception-tolerant: 'if a patient has pneumonia, give him anti-biotics' is a true conditional, even if those patients with pneumonia who are allergic to anti-biotics must on no account get them.¹¹ Indeed one finds subjects who cannot do the task as intended because they view the rule as allowing exceptions¹²

S. If I just looked at that one on its own [7; has A on the back] I would say that it didn't fit the rule, and that I'd have to turn that one [A] over, and if that was

¹¹This observation gives an alternative interpretation of the results of Luria et al. If 'In the Far North, where there is snow, all bears are white.' is interpreted as an exception-tolerant conditional, the subject must be certain that Novaya Zemlya is not an exception. This requires knowledge of the area which the subject quite appropriately denies to possess.

¹²From an experiment reported in [83, Chapter 3], in which the subjects are shown real cards and engage in conversation about their choices with the experimenter.

different [i.e. if there wasn't an even number] then I would say the rule didn't hold.

E. So say you looked at the 7 and you turned it over and you found an A, then?

S. I would have to turn the other cards over ... well it could be just an exception to the rule so I would have to turn over the A.

And this is only the beginning. Careful questioning of subjects shows that they have problems grasping the notion of truth ('turning A and 7 would at most show me the rule is not false; but why is it true?') and of falsity, many interpreting $p \rightarrow q$ is false to mean $p \rightarrow \neg q$ – in which case the choice of the p card only is of course perfectly appropriate. Other subjects have difficulty with the status of the conditional of which they are asked to determine the truth value: 'The experimenter wants to know whether the rule is true or false. But he must know; and would the experimenter (my professor) utter a falsehood?'

The upshot of these results is that experiments in the psychology of reasoning say very little about the rationality or otherwise of human reasoners. Interpreted properly, however, they yield a wealth of information on discourse understanding.

3. LINGUISTICS AND COGNITION

3.1. **A paradox regained.** It has often been claimed that the relations between linguistics and psychology began to be fully appreciated only after the publication of Chomsky's early writings and in particular *Aspects of the Theory of Syntax* [11] in 1965. This is certainly true in view of Chomsky's theoretical mentalism, conceiving of linguistics as dealing ultimately with a system of structures and rules in the language user's mind-brain. However, while theoretical mentalism encourages and perhaps even requires some amount of interaction between the two disciplines, the choice of regarding the study of competence as in principle impervious to the results of experimental research had the effect of sharply separating theories of meaning and grammar from theories of processing. Most commentators agree that the contacts between the two fields have not been as deep and systematic as they should have been, had various obstacles to fruitful reciprocal influence been removed. What they may fail to see is the existence of an inner tension in the very foundation of generative grammar and, as a consequence, the severe inhibiting effect it had on the growth of linguistics within cognitive psychology and neuroscience. Before moving on to the more specific issues concerning linguistics and cognition discussed in this section, it is worth recovering the terms of the paradox directly from Chomsky's text.¹³

One side of the paradox is represented by a number of remarks contained in §1 of the first chapter of *Aspects*, where Chomsky writes:

We thus must make a fundamental distinction between *competence* (the speaker-hearers knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under [...] idealization [...] is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the

¹³Chomsky has entertained different opinions on the issue throughout his career. However, here we choose to focus on those expressed in the 1965 *Aspects* book, not only to avoid shooting at a moving target but also because these have been by and large the most influential. So we identify Chomsky with this specific text rather than with the actual human being and the many positions he has had.

speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. [11, p. 4]

The task of the linguist is to construct a theory of competence as based on performance data, that is on normalized records of linguistic behavior. Chomsky grants that bringing performance to bear on competence is essential to linguistic theorizing. The issue to be settled, which in fact lies at the heart of the paradox, is exactly what counts as linguistic behavior or more precisely what kind of performance data are allowed to shape the theory of competence. Although many generative linguists would contend that it was never a tenet of their research programme to admit data other than native speakers' intuitions, this is apparently not what is suggested by Chomsky's remarks. Indeed, he seems to conceive of a plurality of data types:

Mentalistic linguistics is simply theoretical linguistics that uses performance as data (along with other data, for example, the data provided by introspection) for the determination of competence, the latter being taken as the primary object of its investigations. [11, p. 193]

The evidential basis of linguistics consists of introspective judgments and performance data, which are mentioned here as they were in an important sense different from intuitions. Furthermore, the latter are alluded to as a subsidiary source of evidence and as part of a larger set of data types. Again, the question is what should be considered performance data in this sense and especially whether elicited or experimentally controlled behavior would qualify as such and would thus be allowed to exert some influence on accounts of competence. There are reasons to believe that Chomsky would have answered for the affirmative, the most of important of which has to do with the limits of intuition. In *The Logical Structure of Linguistic Theory* in 1955 he writes:

if one of the basic undefined terms of linguistic theory is 'intuition', and if we define phonemes in this theory as elements which our intuition perceives in a language, then the notion of phoneme is as clear and precise as is 'intuition'. [...] It should be clear, then, why the linguist interested in constructing a general theory of linguistic structure, in justifying given grammars or (to put the matter in its more usual form) in constructing procedures of analysis should try to avoid such notions as 'intuition'. [9, pp. 86-87]

An even more explicit position is expressed in the 1957 book *Syntactic Structures*, where Chomsky suggests that hypotheses on properties of linguistic strings or their constituents should be evaluated on the basis of controlled behavioral tests. Relying on native speaker's judgments or intuitions, he writes:

amounts to asking the informant to do the linguist's work; it replaces an operational test of behavior (such as the pair test) by an informant's judgment about his behavior. The operational tests for linguistic notions may require the informant to respond, but not to express his opinion about his behavior, his judgment about synonymy, about phonemic distinctness, etc. The informant's opinions may be based on all sorts of irrelevant factors. This is an important distinction that must be carefully observed if the operational basis for grammar is not to be trivialized. [10, pp. 8-9]¹⁴

¹⁴The circularity which Chomsky is referring to here is also discussed by Quine in his 1970 paper on linguistic methodology: "We are looking for a criterion of what to count as the real or proper grammar, as over against an

Controlled experimentation is therefore needed in order to overcome the difficulties caused by relying on speaker's intuitions. This presupposes a lucid dismissal of introspective data as an inadequate source of evidence for linguistic theory. So here is one horn of the dilemma: mentalistic linguistics rejects intuitions and requires performance data, including controlled behavioral tests, to determine the theory of competence.

The other side of the paradox is represented by a series of remarks contained in §4 of the first chapter of *Aspects*, where Chomsky questions the nature of the empirical basis of competence theories:

There is, first of all, the question of how one is to obtain information about the speaker-hearer's competence, about his knowledge of the language. Like most facts of interest and importance, this is neither presented for direct observation nor extractable from data by inductive procedures of any known sort. Clearly, the actual data of linguistic performance will provide much evidence for determining the correctness of hypotheses about underlying linguistic structure, along with introspective reports (by the native speaker, or the linguist who has learned the language). [11, p. 18]

Experimental research based on controlled observation and statistical inference is seen as providing facts of no 'interest and importance' and thus rejected as ineffective for the purposes of the theory of competence. Interestingly, intuitions are still treated as they were of lesser importance than performance data. Not for long, however, as Chomsky in the following page takes the decisive step away from psychology:

The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of language to account for masses of evidence that are hardly open to serious question. The problem for the grammarian is to construct a description and, where possible, an explanation for the enormous mass of unquestionable data concerning the linguistic intuition of the native speaker (often, himself); the problem for one concerned with operational procedures is to develop tests that give the correct results and make the relevant distinctions. [...] We may hope that these efforts will converge, but they must obviously converge on the tacit knowledge of the native speaker if they are to be of any significance. [11, pp. 19-20]

The door that seemed to be open for behavioral testing is now closed. The range of data that could affect theories of competence has been narrowed down to intuitions, and more specifically to those of the linguist. The task of the more empirically oriented researcher is to develop tests that will ultimately align with introspective data. The convergence between linguistics and psychology is thus projected forward in time as a desirable outcome not of the joining of efforts, but of their strict segregation. Not only linguistics and the psychology of language are now entirely separate enterprises, but the latter is also required - as a criterion of explanatory adequacy - to provide results which are consistent with the theory of competence as based on the linguist's intuitions. It does not seem inappropriate to observe that, whatever the methodological choices made within linguistics, this is an unacceptable requirement for empirical psychology and neuroscience, which should instead

extensionally equivalent counterfeit. [...] And now the test suggested is that we ask the native the very question we do not understand ourselves: the very question for which we ourselves are seeking a test. We are moving in an oddly warped circle" [71, p. 392]. Root [73] has argued that both Chomsky and Quine, despite the differences, were trying to operationally define linguistic notions, suggesting that Chomsky, at least in his early writings, allowed controlled tests of some sort to affect theories of competence.

be urged to provide solid evidence, no matter which theory or whose intuitions are eventually supported. But what is perhaps more disappointing is Chomsky's disregard toward any form of experimental testing in linguistics:

In any event, at a given stage of investigation, one whose concern is for insight and understanding (rather than for objectivity as a goal in itself) must ask whether or to what extent a wider range and more exact description of phenomena is relevant to solving the problems that he faces. In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. [11, p. 20]

The second horn of the dilemma is thus the following: linguistic theory is based primarily on intuitions of native speakers and does not require controlled experimentation to construct or constrain accounts of competence.

A solution of the paradox, one may argue, amounts to either rejecting mentalism in order to maintain traditional linguistic methodology, in which theories of meaning and grammar are based on intuitions, or adopting a more positive attitude toward the procedures and results of experimental psychology. However, we believe, none of these choices may be sufficient. On the one hand, as we shall see in section 3.3, competence theories as they are usually conceived, that is lacking an explicit link with performance and more precisely with information processing, might be refractory to the results of cognitive psychology and neuroscience, whatever the linguists' attitude toward them. On the other hand, intuitions might still be an undesired source of bias for the theory even if the evidential basis was enlarged so as to include behavioral and other data types. As we shall try to explain below, introspective judgments are unreliable if they are not elicited and recorded in a controlled experiment-like fashion. Intuitions might fail to deliver the desired distinctions and this holds not only for the linguist facing the demands of a theory of competence, but also for the experimenter studying performance. Examples of failures of intuitions in linguistics and empirical neuroscience are provided below and in section 3.4.

3.2. The vagaries of intuition. According to many linguists, in particular generative grammarians and formal semanticists, the intuitions of native speakers constitute the proper empirical basis of theories of competence. However, as we have just seen, the prominent place assigned to intuitions by contemporary linguistics is at odds with the basic tenets of mentalism. If competence is a system of rules which is realized in speakers's mind-brain and if behavior reflects the properties of such system, then constructing a linguistic theory amounts to solving an 'inverse problem', namely inferring the rules of competence from observable performance data. If the (methodological) mentalistic approach to grammar is correct, *any* reliable measure of performance should in principle be allowed to contribute to accounts of competence.

The conflict with mentalism is not the only problem raised by intuitions as an exclusive source of evidence in generative linguistics. Another source of concern is Chomsky's claim that intuitions are not only the starting point of linguistic theorizing, but also the standard to which any grammar should conform:

A grammar can be regarded as a theory of language; it is *descriptively adequate* to the extent that it correctly describes the intrinsic competence of the idealized native speaker. The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be

immediately aware of this) in a substantial and significant class of crucial cases. [11, p. 24]

Even if the tension with mentalism was removed, allowing other data types to influence competence models, and introspective judgments were used only at the outset of linguistic enquiry, intuitions would still pose a number of serious methodological problems. It is not just the role of intuitions in linguistic theorizing that has to be put under scrutiny, but also the philosophical claim that intuitions offer a reliable and privileged window on tacit knowledge in general and linguistic competence in particular.

Several authors have questioned the view of intuitions as providing a transparent access to meaning and grammar. For instance, Jackendoff has observed that the system of rules in the speaker's mind-brain is "deeply unconscious and largely unavailable to introspection" [40, p. 652]. If this is correct, then one should see severe discrepancies between overt linguistic behavior, which reflects the unconscious rules of grammar, and the intuitions or beliefs that speakers have about these rules. This hypothesis has been confirmed by Labov [48] among others, who has collected linguistic evidence on a wide variety of cases in regional American English. One example is the positive use of *anymore* in all sections of the Philadelphia white community, meaning that a situation which was not true some time in the past is now true, roughly equivalent to *nowadays*, as in the following sentence:

(7) Do you know what's a lousy show anymore? Johnny Carson.

Labov interviewed twelve subjects who used the adverb freely and consistently with its vernacular meaning as exemplified in (7) but reported a majority of negative responses when they were asked whether a sentence like (7) is acceptable as well as surprisingly weak intuitions on what the expression signifies in their own vernacular, which contexts are appropriate for its use and what inferences can be drawn from its occurrences.

From different perspectives, converging arguments on the unreliability of intuitions have been proposed. For instance, Marantz has recently emphasized that grammaticality is a technical term defined within linguistic theory [54]: a sound/meaning pair is grammatical or well-formed according to a grammar if and only if the grammar generates or assigns a structural description to the pair such that all relevant grammaticality or well-formedness conditions within the grammar are satisfied. In the passage from *Aspects* quoted above, Chomsky assumes that the structural descriptions assigned by the grammar to sentences can be checked *for correspondence* against native speakers' intuitions. However, native speakers do not have intuitions of grammaticality in this sense, nor they have introspective acquaintance of other properties of strings as defined by a formal grammar. Naive language users might collapse into a single notion of grammaticality different morpho-syntactic, semantic and pragmatic aspects of a sentence. Crucially, they might do so in a way that is beyond control for the linguist. A similar reasoning would apply to intuitive judgments of synonymy versus formal definitions within a semantic theory. Given this discrepancy between the sort of judgments typically provided by native speakers and the intuitions required by linguists, it may seem reckless to regard introspective evidence as being more than a starting point for linguistic enquiry.

As a way out, one may be tempted to argue that a caveat would only apply to naive informants and that intuitions do fair better when the judgments of trained linguists, free from pre-theoretical and confused notions of grammaticality, synonymy and the like are considered. Levelt [49] tested the intuitions of a group of twenty-four trained linguists in a small demonstration experiment. The participants were presented with fourteen examples from their own field's literature, among which:

(8) No American, who was wise, remained in the country.

(9) The giving of the lecture by the man who arrived yesterday assisted us.

None of the linguists judged the ungrammatical item (8) as such and sixteen judged the grammatical item (9) as ungrammatical. Overall, ungrammatical sentences were correctly recognized on average 4.2 times, whereas grammatical items were judged ungrammatical 8.6 times. Levelt warns against taking these results too seriously, but he also observes with some reason that “they are sufficiently disturbing to caution against present day uses of intuition” [49, p. 25].

We could go on providing examples of local failures of intuition, recording the problems that arise with the use of introspective reports in specific judgment tasks. However, we would like to take a more general perspective, considering a type of argument targeted at the nature and scope of intuitions. The argument, which has been proposed by Hintikka [35, 36], starts off with the observation that our intuitions of grammaticality, synonymy, and so on always concern particular sentences or tokens and not entire classes of items or the common grammatical structure they share. Hintikka writes that:

intuition, like sense perception, always deals with particular cases, however representative. [...] But if so, intuition alone cannot yield the general truths: for instance, general theories for which a scientist and a philosopher is presumably searching. Some kind of generalizing process will be needed, be it inductive inference, abduction, or a lucky guess. The intuitions [Chomsky] recommended linguists to start from were intuitions concerning the grammaticality of particular strings of symbols, not concerning general rules of grammar. [36, p. 137-138]

Against Hintikka’s position, one might argue that also paradigmatic variation is a proper object of intuition. On this view, the linguist would be able to grasp the properties of the linguistic structures of interest by constructing a sufficiently large number of particular samples exhibiting that property. This however can be countered with the observation that the supposed ‘intuitions’ about paradigmatic variation are more similar to theory-laden hypotheses: the linguist has to know details of the grammar or the logical structure of the language which are not directly accessible to naive native speakers in order to keep all irrelevant variables constant and systematically manipulate the factor of interest. This leads us to Hintikka’s most important remark. Competence theories do not have built-in devices for deriving abstract syntactic or semantic forms from particular linguistic examples. In any case, inductive generalization from a finite number of cases does not seem the right way to go. On the latter point, Hintikka explains that:

reliance on generalization from particular cases is foreign to the methodological spirit of modern science, which originated by looking for dependencies of different factors in instructive particular cases (often in controlled experimental situations), and by studying these dependences by the same mathematical means as a mathematician uses in studying the interdependences of different ingredients of geometrical figures in analytic geometry. [...] transformational grammarians and other contemporary linguists would do a much better job if, instead of relying on our intuitions about isolated examples, they tried to vary systematically suitable ingredients in some sample sentence and observed how our ‘intuitions’ change as a consequence. Now we can see why such systematic variation is a way of persuading our intuitions to yield general truths (dependence relations) rather than particular cases. [36, p. 135]

If intuitions are to serve as a reliable starting point in linguistic inquiry, they should be proved to have systematic properties. Observing patterns of covariation of introspective judgments and other factors - such as the structure, the content, the context of occurrence of the sentence, the attitude of the speaker and so on - would make the particular examples under consideration instructive and thus effective as part of the empirical basis of linguistic theory. The crucial observation is that, in order to systematically alter the ingredients of sample sentences, the linguist should be able to control these factors in a manner similar to the manipulation of experimental variables in laboratory research. The solution proposed by Hintikka to the problem of intuition points in the direction of infusing linguistic practice with psychological experimentation. The linguist could still start from intuitions, but only the systematic aspects of these as revealed by experimentation and if necessary statistical testing would be preserved and transferred into the theory. Hintikka offers an interesting example in which the imagined goal is to define the meaning of expressions in Montague grammar on the basis of the systematic dependencies between subjects' intuitions and the contexts of occurrence of the expression of interest. In particular, he writes, if the notion of possible world is allowed:

then there is, in principle, no definite limit as to how far your experimentation (construction of ever new situations) can carry you in determining the class of scenarios in which the word does or does not apply. And such a determination will, at least for a Montagovian semanticist, determine the meaning of the word. Indeed, in Montague semantics, the meaning of a term is the function that maps possible worlds on references (extensions) of the appropriate logical type (category). And such functions can, in principle, be identified even more and more fully by systematic experimentation with the references that a person assigns to his terms in different actual or imagined scenarios. [36, p. 146]¹⁵

It is a fair point in favor of introspective reports in a broad sense, however, to add that Hintikka considers thought experiments on a par with genuine experimentation on subjects [36, p. 146]. Instead of eliciting responses from subjects to a number of varying factors, the experimenter imagines herself in such situations. If the variables are controlled with as much care as one would exercise in an experimental setting, then introspection can also reveal the systematic aspects of language and thus contribute to the theory of competence. In this very specific sense, one can indeed say that intuitions deal not only with particular cases but also with paradigmatic variation.

Hintikka's argument can be made more explicit with reference to a number studies investigating the role of the most important of his 'suitable ingredients', namely context. Linguistic and psycholinguistic research has shown that the context in which a sentence is or can be embedded does affect judgments of acceptability. Bolinger [2] observed that sentences, which are judged as semantically implausible when presented in isolation, are regarded as acceptable when embedded in context. Consider for instance the following examples:

¹⁵Although we consider Hintikka's as a passable example of linguistic theorizing based on the co-variation of contextual factors and intuitions, we should also emphasize the problems raised by the notion of meaning in Montague-style semantics. In particular, possible worlds are infinitely many and therefore one cannot even approximate the meaning of an expression using Hintikka's method. In fact, an infinite number of judgments would have to be provided by the subject or by the linguist. For additional criticism of possible worlds semantics, the reader is referred to [90].

- (10) a. It wasn't dark enough to see.
 b. I'm the soup.

For different reasons, these sentences would be judged as semantically deviant: (a) because we normally need light in order to see and (b) because the word *soup* cannot be extended, not even metaphorically, so as to apply to a human being. Now consider the same sentences when embedded in some appropriate context, with (11b) being spoken at a cashier's counter in a restaurant:

- (11) a. I couldn't tell whether Venus was above the horizon. It wasn't dark enough to see.
 b. You've got us confused: you're charging me for the noon special; the man in front of me was the noon special; I'm the soup.

Examples (10a-b) in the appropriate context are perfectly acceptable. Since context has such distinct effects on intuitions, linguistic methodology has to take into account this factor, as Hintikka and Bolinger point out. The same observation applies to experimental research on language processing, as we shall see below discussing an ERP study involving a similar linguistic phenomenon as the one in example (11b). In a similar vein, Levelt et al. [50] demonstrated that judgments of grammaticality vary depending on whether the linguistic material used evoke weak or strong imagery in subjects. On the assumption that subjects perform a judgments task by imagining a context in which the sentence can be uttered, high imagery materials were expected to make the construction a model verifying the sentence easier and faster. Indeed, sentences were judged grammatical more often and more quickly if the verbal material was associated with strong imagery. The conclusion suggested by this study is that the frequency and speed of linguistic judgments systematically vary with variables depending on linguistic context. We take this to be a demonstration that intuitions have also systematic, and not only idiosyncratic, traits. To the extent that the linguist, with the aid of suitable experimentation, is able to exploit the former turning a blind eye to the latter, intuitions can serve as solid starting blocks for linguistic inquiry.

The above remarks on the scope of intuitions within linguistic theory suggest at least that there are reasons to be cautious about them. We would like to make a further step in our discussion considering the use of intuitions outside linguistic theory. The appeal to introspective judgments was never an explicit methodological choice in the psychology and the cognitive neuroscience of language. However, since within linguistics the method was regarded as sound, it was also accepted in language processing research as a means of establishing differences between sentence types or conditions used in actual experiments. The typical procedure is to start with a small number of sample sentences differing with respect to some significant linguistic aspect, the evaluation of which is initially left to the intuition of the experimenter. For instance, to consider one of the first ERP studies on syntactic constraint violations by Osterhout & Holcomb [64], the starting point would be a pair - or a relatively small set of pairs - of sentences containing either an intransitive (12a) or a transitive (12b) verb, using a direct object construction in both cases:

- (12) a. The woman struggled to prepare the meal.
 b. The woman persuaded to answer the door.

Up to this stage the methodology is exactly the same as that of the linguist. However, while the latter directly proceeds with formalizing the requirements of intransitive and transitive verbs with respect to direct objects, the psycholinguist, in order to have sufficient statistical power to test the relevant processing hypothesis in a dedicated experiment, has

to construct a sufficiently large set of sentences with the same structure and properties of (12a-b). In a subsequent step, the items are presented to the subjects while the dependent variables of interest are measured, which in the study of Osterhout & Holcomb were event-related potentials (see section 3.4.1) and grammaticality ratings. To focus on the latter, grammatical sentences like (12a) were judged to be acceptable in 95% of the cases and supposedly ungrammatical items like (12b) in 9% of the cases. One might argue, as a scholarly exercise toward an explanation of the 9% figure, that (12b) does have contexts in which it is both grammatical and semantically acceptable, for instance if it is interpreted as a reduced relative clause (*The woman who was . . .*) and is uttered as an answer to a *who* question, as in the following dialogue:

- (13) A: Who stumbled on the carpet in the hallway?
 B: The woman persuaded to answer the door.

In the context of (13), Osterhout & Holcomb's ungrammatical sentence is in fact acceptable. We have already encountered this phenomenon discussing Bolinger's examples above. Similar contexts of utterance might have been imagined by the subjects who found the sentence acceptable, in a way that escaped the intuitions of the experimenters. Of course, since the 9% figure is relatively small, our observation does not affect the results obtained by Osterhout & Holcomb. What it shows, however, is that subjects' judgments may differ to a variable extent from the intuitions of the experimenter. Moreover, such discrepancies are often beyond control for the experimenter, especially when no or inadequate behavioral tests are carried out. We shall return again to this issue in section 3.4.

Intuitions as used in experimental research are as (un)reliable as they are in linguistics. The experimenter, failing to comply with Hintikka's rule of using thought experiments with systematically varied variables, may treat a grammaticality or synonymy difference between sentences as more rigid than it would be, if a sufficient number of possible contexts of utterance was considered. The experimenter may have strong and correct intuitions on some aspect of the sentences, but may fail to see other salient differences. This kind of situation is not infrequent, but its far reaching consequences are often overlooked. In fact, as we will show using specific examples in section 3.4, inaccurate intuitions at the stage of designing a language processing experiment provide a weak foundation for the study, reduce the predictive power of any theory based on those judgments and seriously bias explanatory accounts.

3.3. Beyond competence and performance. Intuitions are not the only source of concern for a coherently mentalistic approach to language. Jackendoff has pointed out a conflict, which in fact coincides with the paradox as we described it above, between Chomsky's support of mentalism and the choice of maintaining traditional linguistic methodology as based on intuitions and on the competence-performance distinction [39, p. 29]. In the last section we have tried to demarcate the legitimate uses of intuition. This may alleviate the tension, but it is not sufficient to solve the paradox. As we argue below, accepting mentalism requires either a thorough revision or a firm rejection of the dichotomy between competence and performance. Jackendoff has discussed this point, trying to find a more accommodating formulation of the distinction which would allow a natural interplay of linguistic theories and the empirical disciplines investigating language processing. In order to do so, he suggests a return to the original meaning of the distinction:

Chomsky views competence as an idealization abstracted away from the full range of linguistic behavior. As such, it deserves as much consideration as any idealization in science: if it yields interesting generalizations it

is worthwhile. Still, one can make a distinction between ‘soft’ and ‘hard’ idealizations. A ‘soft’ idealization is acknowledged to be a matter of convenience, and one hopes eventually to find a natural way to re-integrate excluded factors. A standard example is the fiction of a frictionless plane in physics, which yields important generalizations about forces and energy. But one aspires eventually to go beyond the idealization and integrate friction into the picture. By contrast, a ‘hard’ idealization denies the need to go beyond itself; in the end it cuts itself off from the possibility of integration into a larger context.

It is my unfortunate impression that, over the years, Chomsky’s articulation of the competence-performance distinction has moved from relatively soft [...] to considerably harder. [39, p. 33]

Jackendoff proposes to adopt a soft distinction between competence and performance, adding a third component to the resulting framework [39, p. 34]. In the new scheme, the theory of competence is the characterization of phonological, syntactic and semantic structures stored and assembled in the mind-brain of speakers in the course of language use. The theory of performance is the description of the use of these structures in the course of language perception and production. The theory of neural instantiation is an account in terms of brain structures and processes of competence and performance. Although Jackendoff’s proposal improves somewhat on Chomsky’s original distinction, it still raises a number of problems that we shall soon review. Let us first delve into how he sees the interactions between competence and performance.

At some point in his book *Foundations of Language*, Jackendoff asks to what extent the logic of competence dictates the logic of processing [39, p. 200]. In syntactocentric architectures, ranging from the standard theory presented by Chomsky in *Aspects* [11] to Government-Binding Theory [12] and the Minimalist Program [13], phrase-structure rules are the only source of combinatoriality in the grammar. Phonology and semantics are seen as interpretive (as opposed to combinatory) systems. More precisely, they take as input a form which has been computed by the syntactic core and interpret it in terms of non-linguistic structures, namely sound and conceptual representations. This entails that phonology and semantics do not add combinatorial structure to the syntactic representation. Rather, a semantic interpretation of a sentence is derived by attaching lexical meanings to the nodes of a syntactic tree - although not to all nodes - and a phonological interpretation results from pairing a grammatical structure with a sound representation. Thus, the logical directionality of the grammar is such that syntax is required to determine phonological and semantic structures. This means that the grammar is neutral with respect to the direction of information flow in comprehension and production. In comprehension, phonological processing precedes syntactic parsing which in turn is followed by semantic interpretation and vice versa in production. Neutrality with respect to the temporal ordering of the steps involved in language comprehension and production is seen as a strength of the theory by generative grammarians. Nevertheless, Jackendoff maintains that syntactocentrism creates an undesirable and unnecessary gap between the logic of competence and the logic of processing. In order to fill this gap, he proposes a novel architecture in which phonology, syntax and semantics are parallel generative systems. The grammar provides interface rules which are part of competence and determine how the combinatory components interact to yield consistent linguistic representations. Interface rules have a prominent role in performance since they act as constraints applied in a parallel and interactive manner during comprehension and production, although in a different temporal order in the two

cases [39, pp. 201-202]. The bottom line of Jackendoff's view, as we understand it, is that interface rules are directly put to use in processing. The following example [39, p. 202] might help clarifying this point:

- (14) a. It's only a PARent, not * a TEACHER.
 b. It's only apPARent, not * REAL.

The initial phrases up to the asterisk mark are acoustically indistinguishable, that is word boundaries cannot be established at the phonological level but only when the subsequent segment is processed. The semantics of *teacher* and *real* constrains the meaning of the preceding clause, disambiguating the acoustic signal. Notice however that the semantic component cannot by itself reject the incompatible phonological structure [39, p. 202]. Rather, there should be interface rules allowing for semantic structures to serve as additional input for the phonological system. Jackendoff's proposal consists in treating lexical items themselves as interfaces between conceptual and phonological structures. Accordingly, the lexicon as a whole is an interface component which, although it contains rules which are essentially different from generative or transformational ones, still forms part of the theory of competence [39, pp. 131, 425]. Jackendoff takes this example as showing that the processor's 'opportunism' is not chaotic, but is constrained by the particular structures available at a given processing stage and by the interactions between these structures through competence rules [39, p. 202]. In short, while in syntactocentric models the grammar is neutral with respect to the logical directionality of competence and performance, in the parallel architecture the logic of competence dictates the logic of processing.

Jackendoff's approach has certainly many merits. In particular, it provides an explicit model in which the different competence components interact with each other in a way that is consistent with the speed and the accuracy of language processing. However, the problem for us - looking for a solution of the paradox - is not to establish that competence determines the state-space available to language users during performance [39, p. 56]. This seems to us the strongest conclusion licensed by Jackendoff's account. Rather, we need to understand whether there is genuine interplay between competence and performance or, in other words, whether the logic of processing can also, to some extent, dictate the logic of competence. Consider for instance Chomsky's observation that speakers of a language are able to produce and comprehend an infinite number of sentences, which is paralleled by the theoretical claim that the productivity of formal generative grammars is infinite. Many commentators have replied that actual speakers, subject to time and memory limitations, can produce and comprehend only a finite number of sentences. This rather unquestionable remark, which clearly pertains to the theory of performance, has important consequences for the theory of competence. In particular, one may wonder whether the use of recursion in formal grammars is adequate at all, since it is on the basis of recursive definitions that the notion of well-formedness covers an infinite number of cases.¹⁶

At least in his early writings, Chomsky would have probably denied that competence theories have anything to learn from processing or performance in general. More recently Minimalists have granted that syntax is optimally adapted to the requirements holding at its interface with other linguistic and non-linguistic domains, such as the sensory-motor and conceptual systems. Even so, they deny what functionalists on the contrary accept, namely that syntax is well-designed for use or optimally adapted to performance [14]. Evidence

¹⁶This line of reasoning leads to alternative approaches to syntax and the syntax-semantics interface such as construction grammar. The reader is referred to Goldberg's works [23, 24] and to the construction grammar website <http://www.constructiongrammar.org>

against optimal adaptation for use is provided, according to Minimalists, by memory limitations, unstable constructions such as garden-path or center embedding sentences, and so on. Here two remarks are in order. The first is that such phenomena do not constitute evidence against optimal adaptation for use. Minimalists seem to be committing what optimality theorists have called a ‘fallacy of perfection’ [57], which consists in equating optimal outcomes - which often are the result of equilibria between different factors - with best possible consequences for one or a few of the factors involved, for instance absence of unstable or ambiguous expressions. The second remark is that, even if we acknowledge that competence is neither perfectly nor optimally adapted for use, this still leaves open the possibility that performance put some constraints, however weak, on competence. This is precisely the question that one needs to raise in order to solve the paradox. We find no hint of an answer in Jackendoff’s parallel architecture and even less so in syntactocentric accounts.

The reason why generative linguistics does not seem capable of issuing this problem is, in our opinion, to be attributed more to how performance is defined than to a rigid view of the competence-performance distinction. As Jackendoff himself admits [39, p. 30], in Chomsky’s original proposal far too many factors were collapsed into performance, namely errors, shifts of attention, memory limitations, processing mechanisms, and so on. Only a very naive view of the problem could justify the claim that a single, relatively compact theory of performance could account for all these factors. It seems more reasonable to assume that several theories, developed with different analytic approaches to the language system, are necessary in order to understand how competence interacts with memory and attention, how errors of different type and origin - for also language disorders give rise to performance failures - are produced, which algorithms are shared by comprehension and production, and so on. Since we think it unlikely that Chomsky and Jackendoff, among many others, could have entertained such a simplistic view of the matter, performance may have been the linguist’s waste paper basket, where everything that does not fit the theory of competence is - irreversibly for Chomsky and temporarily for Jackendoff - thrown. We roughly agree with Jackendoff on the characterization of competence and neural implementation, but we believe that a more appropriate intermediate level should be chosen.

3.3.1. *Marr on the analysis of information processing.* Spivey & Gonzalez-Marquez [77] have remarked that Jackendoff’s reduction of the competence-performance distinction to a soft methodological separation, plus the theory of neural instantiation, resembles Marr’s [55, pp. 24-27] tripartite scheme for the analysis of information processing systems. For our purposes, this is a rather perceptive remark, in that it suggests that Jackendoff’s refinement of Chomsky’s position does constitute a progress, although the trajectory is somehow incomplete. We want to take a further step and adopt Marr’s scheme instead of the competence-performance distinction, at least for the specific problems at hand. This amounts to replacing Jackendoff’s performance theory with an intermediate level of analysis at which algorithms and memory mechanisms are described. This may seem a rather brutal move, restricting the scope of performance to algorithms and thereby leaving out a number of important phenomena which cannot be treated in computational terms. As we hope to show, however, the domain of the theory of performance is now decomposed into more specific problems which are examined at more appropriate levels within Marr’s framework. Let us apply the three-level scheme to language to see how this can be the case.

The first level, which Marr and Poggio [56, 55] call ‘computational theory’,¹⁷ contains a characterization of the data structures which the language system constructs and on which it operates during comprehension and production. The top-level analysis includes also a description of assembly rules and basic constituents stored in memory. For instance, a computational theory for syntax may contain a characterization of phrasal constituents and attachment rules (see below 3.4.4 for an example). A computational theory for semantics may contain a description of the models in which the relevant expressions, translated into a suitable formal language, can be interpreted as well as a definition of valid inference [81] (for an example see 3.4.5). Computational analyses in Marr’s sense and competence theories in Chomsky’s loosely coincide, with one important difference: the notion of an information processing goal is accommodated within the computational theory and thereby within linguistics. The crucial step here is to take the computation of a particular linguistic structure, for instance a parse tree or a model verifying a sentence, as a goal which has to be attained by the system during comprehension and production. Thus, in line with cognitive linguistics, we regard the objects described by the theory of competence as cognitive representations which entertain causal and logical relations with other mental entities. Computational theories are inspired by and testable against behavioral data [56, p. 7], although intuitions are also admitted within Hintikka’s conditions. Analyses at the topmost level are explicitly designed to contribute to comprehensive theories of language processing at all levels of analysis. An important requirement is therefore that they be effectively computable.

The second level contains the theory of algorithms and abstract memory mechanisms. In his 1982 book *Vision* Marr adopts a single level in which only algorithms are included. In a 1976 paper, however, Marr & Poggio [56] split the intermediate level of analysis into two, with the result that the theory of memory mechanisms is independent from algorithms and mediates their implementation into the physical system. We will adopt a more flexible position here, assuming that algorithms and memory mechanisms can be described at the same level of analysis or at least that their analyses can be coupled. We will give priority to algorithms, whose description is in any case mandatory, and discuss memory mechanisms only when necessary. One reason for this choice is that the same techniques can be used to test algorithms and memory mechanisms. Behavioral measures like reading or reaction times and physiological data such as event-related potentials (see section 3.4.1) can be used for this purpose since they typically allow reliable inferences concerning the time course and the complexity of cognitive processes, which in turn can be used to test hypotheses on whether computation is serial or parallel, immediate or delayed, and so on.

The third level contains the theories of neural implementation. Contrary to Marr and Poggio and Jackendoff, we use and emphasize the plural here because analyses of brain systems carried out at different levels of granularity are required. Coarse-grained analyses may amount to a description of the spatial and temporal dynamics of brain processes as revealed by neuroimaging techniques such as EEG, MEG and fMRI, possibly integrated by lesion data. In addition, these accounts may include statistical maps displaying the functional and anatomical connectivity between the different systems involved in language processing as well as between language and other domains, such as the visual, auditory and motor systems. Finer-grained analyses may contain descriptions of the properties, the connections and the firing patterns of the neurons forming the structures under exam. It is

¹⁷This is a rather infelicitous terminological choice, since computational theories in Marr’s sense concern more the goals of computations rather than the actual computations themselves. Informational level of analysis would perhaps be a better term.

important to reiterate that the motivation for having a relatively independent level of neural implementation is not that of providing empirical data which are used to test computational or algorithmic theories. Indeed, in Marr's scheme the relation between theory and data is not articulated along the vertical axis of levels of analysis. Quite the contrary, the theory of physical implementation contains mathematical models as well as hypotheses to be verified on a specific set of data types, just as computational theories are evaluated on the basis of overt or controlled behavior and algorithms are tested on reaction or reading times and physiological data. Ideally, the result would be a unified scheme which allows to specify in a vertical manner the relations between formal theories of competence, processing algorithms and models of neural computation and in an horizontal manner, for each of these level, the empirical context of application.

The issue that should now be addressed is whether Marr's framework solves the problems associated with Chomsky's and Jackendoff's competence-performance distinction. One of the most significant improvements, we believe, is that the different factors which were collapsed into performance can now be properly understood in their specific features. The working memory resources required by the relevant algorithms can be examined at the intermediate level of analysis. The study of shifts of attention may begin with a general description of the attentional demands of language use and proceed with an account of the algorithms allocating attentional resources to the different sub-tasks involved. A study of the on-line interactions between language and attention may also be useful to determine the conditions in which the allocation mechanisms fail. The theory of processing - itself a large waste paper basket - has to do essentially with modeling and testing realistic algorithms for a well-specified computational theory. Errors might arise from the interplay of different factors, depending on the nature of the algorithms, the available computational resources and the normal or impaired functioning of the relevant brain system. The other major problem, namely competence being shielded from performance, is now open to discussion: the range of admissible linguistic structures may be narrowed down to some extent when we consider only those that can be computed by realistic algorithms, memory mechanisms and neural architectures, that is those which are globally consistent with theories at other levels of analysis. This gives also a tentative answer to the question to what extent the logic of processing dictates the logic of competence, since it points to at least one well-defined bottom-up constraint, namely effective computability.

Our appeal to Marr's framework was motivated by the difficulties arising with the traditional view, which conflated a number of heterogeneous factors into performance and separated the theory of grammar from the theory of processing. The three-level scheme makes the relations between competence theories and empirical data clearer, through a study of processing algorithms and their physical implementation. In a sense, our account is a version of the classical functionalist view of the relations between mind and brain as applied to the case of language. It seems possible to ask, especially from a neuroscience-oriented standpoint [95], whether functionalist frameworks raise more problems than they solve. First, if analyses can be carried out independently at each level or if the levels are only loosely related, as Marr and Poggio have emphasized in several occasions [55, p. 25] [56, p. 2], and multiple solutions at each level are possible, then the computational and algorithmic theories are neutral with respect to the physical implementation. This seems to make competence theories irrelevant for cognitive neuroscience and vice versa. Second, if computational accounts cannot be derived from the observation of the physical machinery at work, as Marr has claimed commenting on Barlow's first dogma [55, p. 12-15], then it is again hard to see how low-level data could have any effect on competence theories. This

might open a gap between linguistics and cognitive neuroscience, similar to the one caused by the paradox described above.

However, these remarks don't do justice to the moderate version of functionalism we are adopting here, which on the contrary may have desirable effects on the relationships between linguistics and neuroscience. Consider computational models in psychology and neuroscience, which usually consist of neural network models of cognitive processes rendered as computer programs, simulating the behavior of actual system. Marr has observed [55, pp. 27-29] that cognitive models in this sense fail to account for what he calls 'the importance of computational theory'. Without a precise description of the overall function to be computed, the design of the algorithms can only proceed in a bottom-up fashion, adjusting the parameters of the network until it yields the desired results. Especially as far as language is concerned, this approach might provide models which compute efficiently some specific function, for instance the degree of relatedness between the meaning of a word and its semantic context, but fail to account for how the computation of the function contributes to solving the actual information processing problem, for example interpreting a sentence in a discourse model. In order to correctly integrate the particular functions into the larger computational problem, it is crucial to know what kind of problem the latter is, for instance whether it amounts to computing the value of a function rather than to solving a constraint-satisfaction problem, and so on. What is required is the top-most computational level of analysis, independent from the algorithms, in which such issues are settled. Marr's functionalism could infuse cognitive neuroscience with a deeper and more articulated notion of computational modeling, starting from a simple and clean mathematical account of the overall computation, subsequently decomposed into sub-goals and implemented into a set of algorithms and memory mechanisms effectively carrying out the computation. Moreover, the predictive and explanatory powers of a processing model might increase when an abstract computational theory is specified. We provide some examples suggesting that this is indeed the case in section 3.4 below.

3.4. Experiments on semantic processing. In this section we review four recent studies on semantic processing, with two main purposes. The first is to exemplify the claims about intuitions and levels of analysis made above. The second is to introduce and apply the models of syntactic and semantic binding which shall be motivated and discussed from a wider perspective in section 4. The first two studies - Kim & Osterhout [43] and Kuperberg et al. [45] - provide opposite accounts of very similar experimental results. These impasses are not infrequent in language processing research, but there is a sense in which they can be instructive. When no decision on the correct explanation can be made, appealing to the poverty of empirical data may seem premature if no clear computational model licensing precise predictions is proposed. Controversy in language processing research is more likely to be settled when computational and algorithmic analyses, testable against empirical data, are provided. The other two studies - Nieuwland & van Berkum [63] and Münte et al. [62] - provide accounts of experimental results which, upon a closer examination, have at least equally strong competitors based on explicit computational models of syntactic and semantic processing. We shall sketch these alternative analyses and speculate on their explanatory and predictive powers. Let us first introduce the technique with which these experiments were carried out, namely event-related potentials (ERPs).

3.4.1. Event-related potentials. Since the late 1920s it has been possible to record the electrical currents generated by the brain by placing a number of electrodes on the scalp, amplifying the signal and plotting the observed voltage changes as a function of time. The

resulting electroencephalography (EEG) is thought to reflect the activity of a large number of cortical sources. Only a small portion of the electrical potentials measured at the scalp is evoked by the relevant sensory, motor or cognitive event, for instance a stimulus presented to the subject or a response she is required to make. There exist different techniques to extract event-related potentials from the underlying EEG. For instance, it is possible to reconstruct ERPs from single-trial EEG segments, that is from the raw signal recorded after a single stimulatory event [42, 70]. However, averaging over a relatively large number of trials of the same type is the most widely used approach to obtaining ERPs. The assumption here is that noise, defined as the activity which is not evoked by the cognitive process of interest, is randomly distributed in terms of polarity, latency (with respect to the onset of the stimulus or the response) and amplitude in the EEG trace. As a consequence, averaging will tend to reduce the noise and reveal the ERP signal.¹⁸ ERPs seem particularly suited for the study of information processing [17] as they provide an excellent window on the temporal dynamics of neural activity. Furthermore, ERPs have a number of advantages over cumulative measures like eye movements, reading and reaction times in that they provide qualitative information about the processing stages affected by the experimental manipulation [52].

Event-related potentials have proved useful to address a number of issues concerning the relative time course and the complexity of phonological, syntactic and semantic binding operations. Kutas & Hillyard [46] conducted the first ERP experiment in which linguistic factors were manipulated, in the specific case the semantic plausibility of a word in a given sentence context:

- (15) a. The officer shot the man with a gun.
 b. The officer shot the man with a **moon**.

Compared to the control word *gun*, the anomalous item *moon* resulted in a larger negative shift starting around 300 ms following word onset, peaking at 400 ms and lasting until 500 ms. The component, which was called N400 because of its polarity and peak latency, was not elicited by other unexpected events such as variations in the physical properties of stimuli, for instance the size of the words presented on the subject's screen. Larger N400s are elicited also by semantically plausible words which are nevertheless judged by subjects as less preferred in a given sentence context [47, 30], like *pocket* below:

- (16) a. The girl put the sweet in her mouth after the lesson.
 b. The girl put the sweet in her **pocket** after the lesson.

Finally, larger N400s compared to the baseline¹⁹ are evoked by lexical items which provide information conflicting with the preceding discourse [87, 88] or with world knowledge [33]. These results have led researchers to interpret the N400 as an index of the difficulty of integrating word meaning into a semantic representation of the unfolding sentence or discourse.²⁰ Semantics-related sustained negative shifts - different from the N400 - are elicited by constructions which are considered taxing on working memory such as referentially ambiguous nouns [86, 85] and temporal connectives [62] (see section 3.4.5). Research on the neural correlates of semantic binding operations has produced a considerable amount of evidence. Nevertheless, computational and algorithmic accounts of semantic composition,

¹⁸For details, see Luck's book on ERPs [52], Chapters 1 and 4 in particular.

¹⁹Every word elicits an N400

²⁰There are however recent studies showing that larger N400s are also elicited by co-speech gestures which do not match with the semantic content of the accompanying sentence. This suggests that the N400 indexes also cross-modal semantic integration. See the study of Willems et al. [96] for experimental data and a discussion of the issue.

embracing linguistic theories, behavioral and electrophysiological data, seem still beyond the horizon of present research.

An ERP component associated with syntactic processing was reported independently by Osterhout & Holcomb [64] and Hagoort et al. [32] in sentences violating grammatical constraints:

- (17) a. The spoilt child is throwing the toy on the ground.
 b. The spoilt child **are** throwing the toy on the ground.

Compared to the control item, the incorrect verb form *are* elicited a larger positive shift starting around 500 ms following the onset of the violation and lasting for about 500 ms. The component was called P600 [64] on the basis of its polarity and peak latency or SPS (Syntactic Positive Shift) [32]. Larger P600/SPS components are also elicited by grammatically anomalous sentences in which semantic and pragmatic factors have been removed [30]:

- (18) a. The boiled watering-can smokes the telephone in the cat.
 b. The boiled watering-can **smoke** the telephone in the cat.

A P600/SPS was also evoked in temporarily syntactically ambiguous sentences by the word which introduces the only parsing option available, as in (19b) compared to the unambiguous sentence (19a) [31]:

- (19) a. The sheriff saw the indian, and the cowboy noticed the horses in the bushes.
 b. The sheriff saw the indian and the cowboy **noticed** the horses in the bushes.

While in (19a) *the cowboy* can only be the subject NP of the VP starting with the verb *noticed*, in (19b) it could also be a direct object of the preceding VP, that is until the following VP is processed.

Research on language processing has revealed another family of ERP components, namely (early) left-anterior negativities or (E)LANs (see [31] for a brief overview), associated with morpho-syntactic violations and with filler-gap dependencies (for a recent study, see [65]). Efforts toward a computational framework accounting for syntax-related ERP effects have been made by Hagoort [27] using a lexicalist parser elaborated by Vosse & Kempen [94], as we shall explain below. We now turn to the four ERP studies we have chosen to review.

3.4.2. *Semantic attraction*. Kim & Osterhout [43] examined the hypothesis, common to a number of psycholinguistic models, that syntax is informationally encapsulated, that is it operates without interaction with the other components of the language system [20].²¹ According to these models, when the syntactic structure is unambiguous, the parser will drive the analysis of the linguistic material ignoring input from other domains, no matter which alternative structure - for instance based on more plausible or more frequent theta role assignments - is suggested by semantic cues. Semantics might influence processing only when syntactic structure is ambiguous. In the first ERP experiment reported in [43], Kim & Osterhout presented subjects with violation sentences (20a), passive controls (20b), and active controls (20c):

- (20) a. The hearty meal was **devouring** the kids.
 b. The hearty meal was devoured with gusto.
 c. The hungry boys were devouring the plate of cookies.

²¹Informational encapsulation should not be confused with domain specificity, which makes a different and possibly weaker claim, namely that syntax and semantics are supported by qualitatively distinct neural processes, as revealed by the differences between the P600/SPS and the N400.

Violations were such that the syntax unambiguously assigns the Agent role to the noun *meal* and the Theme role to *kids*.²² Semantic cues, however, would suggest a more plausible theta role assignment, in which the inanimate *meal* is Theme and the animate *kids* is the Agent. Kim & Osterhout reasoned that, if unambiguous syntactic cues drive combinatory processing ignoring semantic input, (20a) sentences should be perceived as semantically anomalous and thus elicit an N400. The thematic roles assigned by the syntax would cause the word *devouring* to be treated as semantically inappropriate compared, for example, to verbs which are consistent with the inanimacy of *meal* such as *satisfying*. Alternatively, if semantics can influence processing when the syntax is unambiguous but would assign implausible thematic roles to nouns, (20a) sentences should be perceived as syntactically ambiguous and thus elicit a P600/SPS. The semantically appealing interpretation of *kids* as Agent and *meal* as Theme would cause *devouring* to be regarded as grammatically inappropriate compared to the past participle form *devoured*. The elicitation of an N400 or a P600/SPS at the verb would be taken by Kim & Osterhout as evidence that syntax or semantics respectively guide combinatory processing. Contrary to the predictions licensed by models assuming the autonomy of syntax, violation sentences evoked a clear P600/SPS compared to both (20b) and (20c) items. This points at a semantic attraction effect toward the most plausible theta role assignment, such that the system processed (20a) as syntactically anomalous even though the sentence is grammatically well-formed.

Against their own account of the data, Kim & Osterhout observe that the P600/SPS might not be due to a semantic attraction effect, resulting in a perceived syntactic anomaly, but to an outright syntactic violation. In (20a) the inanimacy of the subject noun *meal* might have predisposed the system to assume a passive analysis for the sentence according to which the passive participle verb *devoured* should be expected. The upcoming verb *devouring* might have been perceived to be in the wrong grammatical form, resulting in a P600/SPS. To counter this objection, a second experiment compared (21a) and (21b) as above to a non-attraction violation (21c) in which the initial noun phrase was also inanimate:

- (21) a. The hearty meal was **devouring** the kids.
 b. The hearty meal was devoured with gusto.
 c. The dusty tabletops were devouring with gusto.

The comparison between conditions (21a) and (21b) resulted again in a P600/SPS. In (21c) the subject noun is inanimate, so if the system was predisposed to a passive analysis, revised when *devouring* is processed, a P600/SPS would result from perceiving the wrong verb form. On the contrary, an N400 effect was elicited by the verb in (21c). Since the semantic relatedness of *tabletops* and *devouring* is relatively weak, there is no semantic attraction effect that could favor an interpretation in which the tabletops were in fact devoured. As a consequence, semantic unrelatedness results in an outright semantic violation as the N400 indicates. According to the authors, this is additional evidence that the P600/SPS observed in the critical comparisons involving (21a) was not due to the inanimacy of the subject noun, but was an effect of the semantic relatedness of the words, suggesting a different thematic role assignment than the one unambiguously signaled by

²²On most linguistic accounts theta role theory is a separate component of the grammar and lies at the interface of syntax and semantics. Kim & Osterhout seem to take theta role structure as part of the syntax proper and this might generate some confusion for the more linguistically-oriented reader. We believe that this loose and perhaps improper notion of theta structure does not affect the hypotheses and the results presented in [43]. What it does imply, however, is that the standard association of the P600/SPS with syntax has to be taken with a grain of salt. A similar caveat applies to the study of Kuperberg et al. discussed below.

syntactic cues. In conclusion, inanimacy is not sufficient to elicit the P600/SPS while semantic attraction is necessary in order to evoke the effect.

One problem with the second experiment is that Kim & Osterhout do demonstrate that semantic attraction is required in order to give rise to a perceived grammatical violation and thus to a P600/SPS, but the result is restricted to the case in which the subject noun is inanimate. Only in this particular case, the experiment succeeds in showing that inanimacy is not sufficient to give rise to the effect and that semantic attraction is necessary. However, it may also be that semantic attraction, when the initial noun is animate, is not sufficient either and that therefore inanimacy and semantic attraction are necessary but not sufficient antecedent conditions for the elicitation of the P600/SPS. The data are explained also by an account in which the perceived syntactic anomaly is the consequence of an interaction between the inanimacy of the subject noun and the semantic relatedness between lexical items resulting in the attraction effect. An experiment that could contribute to settling the issue would be designed crossing the factors animacy and semantic attraction, obtaining the following four sentence types, possibly to be added to the controls used by Kim & Osterhout:

- (22) a. The hearty meal was **devouring** the hungry kids. (ANIM–, ATTR+)
 b. The dusty tabletop was devouring the hungry kids. (ANIM–, ATTR–)
 c. The hungry kids were **satisfying** the hearty meal. (ANIM+, ATTR+)
 d. The hungry kids were satisfying the dusty tabletop. (ANIM+, ATTR–)

The results reported by Kim & Osterhout are consistent with an interaction of animacy and attraction, since (21a) sentences elicit a P600/SPS and (21b) sentences an N400 compared to controls. The hypothesis that semantic attraction is necessary and sufficient while animacy is neither necessary nor sufficient to give rise to a perceived syntactic anomaly - which we assume is what Kim & Osterhout wanted to prove - would be supported only if a main effect of attraction was found and more precisely if also (21c) elicited a P600/SPS and (21d) an N400 compared to controls. However, on the basis of a simple linguistic consideration, we suspect that a somewhat different pattern of results would be observed for (21c) and (21d). Compared to inanimate subjects, animate subjects place less constraints on possible theta role assignments. Inanimate subjects like *hungry kids* can indeed act as the Theme of *satisfying*. As a result, a violation may be detected downstream of the verb, when the final noun phrase is processed. We will return on the conclusion of Kim & Osterhout after having considered one of the predecessors of their study.

3.4.3. *Conceptual constraints*. Kuperberg et al. [45] investigated the ERP correlates of violating two types of conceptual constraints within sentences. The first are violations of thematic role assignments between nouns and verbs in which contextual coherence is nevertheless preserved. The second are violations of contextual plausibility preserving thematic roles. Control items (23a), thematic role violations (23b) and pragmatic violations (23c), as the authors call them, were presented to subjects:

- (23) a. For breakfast the boys would only eat toast and jam.
 b. For breakfast the eggs would only **eat** toast and jam.
 c. For breakfast the boys would only **bury** toast and jam.

An N400 was elicited by the verb in pragmatic violations which was larger than the corresponding component evoked by verbs in control sentences and in thematic role violations. The verb in theta role violations resulted in a larger P600/SPS compared to the other two conditions.

As Kuperberg et al. acknowledge, the interpretation of the attenuated N400 in thematic

role violations compared to pragmatic violations is problematic. One possibility is that the amplitude of the N400 is inversely correlated with the degree of semantic relatedness between the lexical items in the sentence: the weaker the semantic relation, the larger the N400 amplitude. The weak association between the items *breakfast* and *bury* in (23c) compared to *breakfast*, *eggs* and *eat* in (23b) would explain the N400 amplitude difference. However, this account factors out the possible effects of the anomaly contained in (23b), whatever its exact nature. This could be justified either by providing a processing model in which the relevant algorithms compute degrees of semantic relatedness in an independent fashion with respect to the assignment of thematic roles or by showing experimentally that the amplitude of the N400 correlates with semantic relatedness irrespective of the presence or absence of a thematic role violation. Both options are open to inquiry, but none of them is pursued by Kuperberg et al.

The second possibility considered by Kuperberg et al. takes into account the impact of the thematic role violation on semantic integration. On this view, the amplitude of the N400 is attenuated in (23b) because the detection of a theta role violation disrupted semantic processing. The suggestion here is that thematic role assignment precedes and feeds into semantic integration. In order to support this claim, a model should be provided in which the algorithms computing theta role structure and semantic interpretations are spelled out in some detail. In particular, it should be shown that thematic roles can be assigned on the basis of syntactic cues only, effectively confining lexical processing to a subsequent step. However, Kuperberg et al. [45, p. 127] claim that theta role assignment might require lexical processing to some extent, in particular animacy information. The paradoxical outcome is that, while the words *eggs* and *eat* must be processed - as an inanimate entity and as an action verb respectively - before a thematic role can be assigned to *eggs* and thus before a theta role violation is detected when *eat* is encountered, the N400 elicited by *eat* was attenuated because lexical processing was disrupted by the detection of the violation. In a nutshell, the argument is circular because, if processing of the word *eat* is abandoned when the theta role violation is detected, it cannot be at the same time required by the assignment of a thematic role to *eggs* upon which the violation depends. This paradox may be only apparent, but clearly an analysis at the level of algorithms is needed in order to determine what is required at each computational stage and which process, if any, is disrupted or abandoned. The price of avoiding a formal analysis is losing grip on the functional significance of the N400.

Similar problems arise when an explanation for the P600/SPS is sought. Kuperberg et al. tentatively accounted for the attenuated N400 by hypothesizing that lexical integration was disrupted by the thematic role violation. Yet, they take the opposite stance in their attempt to explain the P600/SPS, where they assume that lexico-semantic animacy information is required to construct a thematic structure. We refrain from insisting on this inconsistency and we focus instead on the details of the explanation of the P600/SPS. The argument used is that, while processing the noun *eggs*, a probable theta role assignment is computed using animacy information, with *eggs* being the Theme of some as yet unspecified action. When the verb *eat* signals that *eggs* is in fact the Agent, the newly computed thematic structure conflicts with preceding one, resulting in a P600/SPS. On this account, first a semantically attractive interpretation is pursued, assigning the Theme role to *eggs*, and subsequently the reading forced by the unambiguous syntactic cues is adopted, in which the Agent role is assigned to *eggs*. Let us now get back to study of Kim & Osterhout and consider again the P600/SPS elicited by (20a) sentences:

- (20) a. The hearty meal was **devouring** the kids.

According to the semantic attraction hypothesis of Kim & Osterhout, when *devouring* is processed, the syntactic cues would unambiguously signal that *meal* is Agent. However, semantic attraction between the lexical items - possibly interacting with animacy - would very rapidly suggest that *meal* is a better Theme for *devouring*, resulting in a perceived grammatical anomaly at the verb. According to the account of Kuperberg and colleagues, lexico-semantic and animacy information would initially suggest the most probable Theme role for *meal*. However, when the verb is processed, the unambiguous syntactic cues would force the parser to assume that *meal* is the Agent, resulting in a conflict with the preceding assignment. In short, the exact same effect can be explained by two opposite processing mechanisms. In one account, syntactic analysis based on unambiguous cues is overridden by semantic attraction and thus by a more plausible thematic role assignment. In the other, the most probable thematic role assignment given animacy information is overridden by a syntactic analysis based on unambiguous cues. Similar antithetical explanations could also be sketched for the P600/SPS observed by Kuperberg et al. Further experimentation along the lines outlined above with sentences (22a-d) might help to disentangle the contributions of animacy and semantic attraction in eliciting the P600/SPS. However, as long as explanations are not based on algorithms specifying the mutual requirements of syntax, animacy and theta role assignment, the data might remain silent as to which hypothesis - semantic attraction or syntactic forcing - is correct. As a matter of fact, both accounts seem compatible with the robust finding that the P600/SPS requires a violation in syntactic structure as an antecedent condition. In this case, we are inclined to conclude that it is more the poverty of the proposed models than the poverty of the data which should be held responsible of the impasse. We now turn to the other two studies committing ourselves to proposing alternative computational analyses of the observed phenomena.

3.4.4. *Semantic illusions.* Nieuwland & van Berkum [63] investigated the ERP effects of coherence breaks in discourse. They presented subjects with ministories of four sentences (24a) in which a man and a woman engaged in a conversation about an inanimate object. In each of these sentences the man and the object were mentioned once. In the fifth sentence, however, the woman either continued the conversation with the man as in (24b) or with the inanimate object as in (24c). In the latter case, the noun referring to man in the control condition was replaced by the one referring to the object in the critical one:

- (24) a. A tourist wanted to bring his huge suitcase onto the airplane. However, because the suitcase was so heavy, the woman behind the check-in counter decided to charge the tourist extra. In response, the tourist opened his suitcase and threw some stuff out. So now, the suitcase of the resourceful tourist weighed less than the maximum twenty kilos.
 b. Next, the woman told the tourist that she thought he looked really trendy.
 c. Next, the woman told the **suitcase** that she thought he looked really trendy.

The appearance of the word *suitcase* in (24c) marks a strong coherence break in the story. Accordingly, an N400 would be expected at the critical word. However, the N400 elicited by *suitcase* was not larger than the corresponding component evoked by *tourist*. Instead, a larger P600/SPS was elicited by critical items compared to controls. Presenting (24b) and (24c) sentences without previous context did result in a larger N400 for *suitcase* compared to *tourist*.

The account of Nieuwland & van Berkum is based on a number of inferences from the observed data to the absence of the cognitive process of interest. First, they seem to take a null effect of condition in the N400 time window as evidence that the replacement of

tourist with *suitcase* did not result in differential processing. The argument is that, since even very subtle differences in lexical integration result in larger N400s, the absence of a modulation of the component indicates that the replacement went undetected. Obviously, this claim needs to be supported by behavioral data showing that subjects failed to notice the coherence break. Without behavioral corroboration, one might also argue speculate that the anomaly was detected and that this resulted in a significant ERP effect - the P600/SPS. The problem here is not only a missing piece of behavioral evidence, but also the lack of a computational model accommodating the N400 and the P600/SPS in a way that the former is elicited by the immediate detection of a coherence break or an inappropriate exchange of referents - which, notice, is not the same as lexical integration - while the latter effect cannot be elicited by such process. Without this sort of theory it is a matter of preference whether to accept the absence of an N400 as evidence that the coherence break went temporarily unnoticed or to take the presence of a P600/SPS as evidence that a coherence break was in fact detected. Similarly, without a top-level analysis, it remains unclear whether and how exactly a ministory in which discourse referents are introduced and mentioned several times can induce or facilitate semantic illusions as, according to the authors, would be suggested by the presence of an N400 when critical sentences are presented in isolation.

Not surprisingly, Nieuwland & van Berkum are unable to explain the P600/SPS. One hypothesis they consider is that “although the immediate *symptom* that signals a comprehension problem might be semantic, the system might in these cases ‘put the blame on syntax’ (e.g., consider the possibility of an incorrect phrase ordering)” [63, p. 699]. An alternative account would be to assume that “the P600/SPS more generally indexes a monitoring process that checks upon the veridicality of ones sentence perception” [63, p. 699]. Whatever the exact meaning of these possibilities, a few points are clear. If, following the first hypothesis, the P600/SPS reflects a syntactic response to a semantic comprehension problem, then the detection of the coherence break must have occurred before the syntactic response, otherwise either there is a timing paradox or there is nothing which the P600/SPS constitutes a response to. The only possible conclusion is that the neural correlate of such detection process simply did not show up in the ERP, which is at odds with their claim that, in fact, the anomaly went unnoticed, that is no detection process took place before the P600/SPS. Alternatively, if the P600/SPS is not interpreted as a brain signature of syntactic processing, but becomes a ‘monitoring process’ or, as Nieuwland & van Berkum suggest [63, p. 699], an index of a “delayed and more careful interpretive process that subsequently overrides [the initial] incomplete, illusory interpretation”, then the authors should be prepared to accept the more parsimonious interpretation - equally distant from standard accounts of the positive shift and presumably equally implausible - of the P600/SPS as reflecting the detection of a coherence break. But even accepting in toto the explanation of Nieuwland & van Berkum, it remains doubtful whether the P600/SPS can count as a ‘delayed and more careful’ response to the anomaly. There is of course no criterion as to what counts as a delayed response, but probably the few hundreds milliseconds separating the N400 - which, according to the authors, would be the appropriate response, thus not a delayed one - from the P600/SPS might not be enough.

The main problem with the study of Nieuwland & van Berkum is that the phenomenon investigated, namely the replacement of a noun with another one, both having accessible referents in the discourse model, is not clearly distinguished from lexical integration, and thus it is open to question whether an N400 should be expected at all. Similarly, the authors do not propose a model in which noun replacement is clearly distinguished from syntactic binding, which agains makes it uncertain whether a P600/SPS should not be expected.

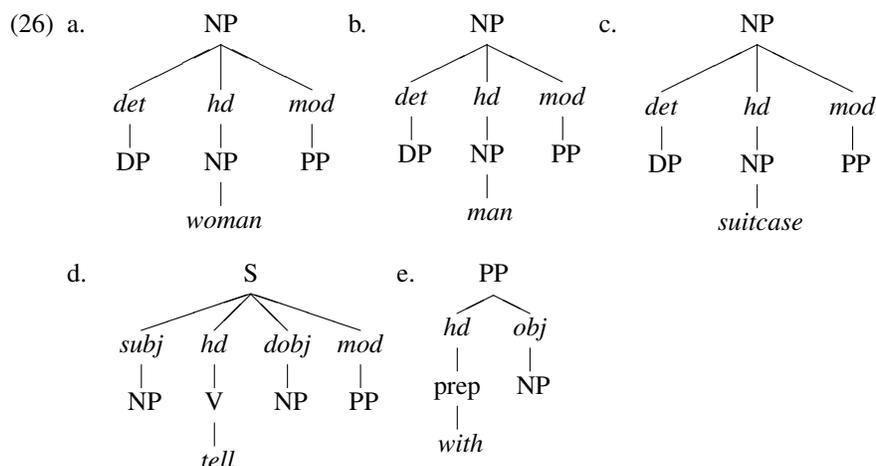
In an important sense, it is the absence of a computational theory in Marr's sense which throws explanations into disarray. Below we try to outline an alternative computational analysis of the P600/SPS observed by Nieuwland & van Berkum.

Sentences like (1c) above can be seen as instances of metonymy or coercion, where the noun *suitcase*, normally referring to an object belonging to the man, is in this case used to refer to the man himself. Nieuwland & van Berkum dismiss these constructions because they "derive much of their attractiveness as stylistic or humoristic devices from the very fact that they are less expected" [63, p. 699]. However, this is a clear example of how the narrow intuitions of the experimenters concerning the appropriate contexts of use of an expression can lead to a false generalization. As a matter of fact, coercion is used in many more cases than the humorous or stylistically marked discourses mentioned by Nieuwland & van Berkum. The following examples are taken from Jackendoff [39, p. 388]:

- (25) a. [One waitress to the another:]
 The ham sandwich in the corner wants some more coffee.
 b. Plato is on the top shelf next to Russell.
 c. John got a dent in his left fender.

Example (25a), far from being humorous or indicating a particular stylistical choice, might on the contrary make use of coercion to deal effectively with a situation - for instance, a busy restaurant - in which utterances should be maximally brief and informative. Sentences like (25b), used in the context of a bookstore or a library as an indication on where to find the desired books, would be probably preferred with respect to the lengthier description *The works of Plato are on the top shelf next to those of Russell*. Anyone with some experience of libraries or bookstores would agree that such shorthands are in fact relatively frequent. Similar considerations hold for (25c), which uses the proper name and the personal pronoun to construct an efficient shorthand for an otherwise lengthier sentence, without being either humorous or stylistically marked. The crucial factor enabling coercion in all these cases is context. If the information available to the hearer makes clear which entities the speaker is referring to, then coerced constructions are permissible and, when pragmatically more efficient, also preferred. Context is thus the first ingredient of our alternative account.

The second ingredient is borrowed from syntactic theory, in particular from a lexicalist parser originally developed by Vosse & Kempen [94] and subsequently used by Hagoort [27] as a computational model of syntactic unification. In this model, syntactic information is stored in the mental lexicon as three-tiered unordered trees specifying the admissible syntactic environment of a particular lexical item. To consider the critical sentence (24c), when the words *woman*, *man*, *suitcase* and *told* are processed, the following syntactic frames become part of the unification space:



In all four cases, the top layer of the frame is constituted by a single root node containing phrasal information (NP, S, etc.), to which unordered functional nodes representing the syntactic environment of the word are connected. The nodes in this second layer depend on the phrasal constituent occupying the root node. For instance, an S node will have four functional nodes specifying its syntactic environment, namely subject and modifier. An NP node will have three functional nodes, namely determiner, head and modifier. The third layer contains again phrasal nodes, called foot nodes, to which the root node of another lexical item can be attached. Syntactic unification consists first in checking agreement and other features, including lexico-semantic properties of words like animacy, and subsequently, if the check is passed, in linking up frames with identical root and foot nodes. We cannot enter the details of the theory here, but in short the unification mechanism is based on competition between alternative attachments and lateral inhibition to suppress discarded options. Hagoort proposes that the P600/SPS is related to the time required to establish attachments of sufficient strength. Such temporal variable is affected by syntactic ambiguity (i.e. the amount of competition at a given stage), syntactic complexity, semantic influences and failed binding attempts as in the case of syntactic violations.

The following computational analysis can be derived from the model sketched above. The three NP frames (26a-c) are introduced in the unification space when the ministry (24a) is processed. Sentences (24b) and (24c) introduce the additional frame (26d) of the verb *told*. In the control case (24b), the NP root node of *woman* is bound to the identical foot node of *told* as a subject.²³ When *man* is encountered, successful animacy checking allows its root node to be bound with the remaining NP foot node of the verb as a direct object. Unification is not as smooth in the critical (24c) case. Again, the root node of *woman* binds to the subject NP foot node of *told*. However, when *suitcase* is encountered, the animacy restriction will prevent its root node from binding with the identical direct object NP node of *told*. Instead, the NP root node of *suitcase* will be bound to the foot NP node of the preposition *with* in (26e), whose root node is in turn bound to the PP foot node of *man*. In other words, the sentence is treated by the processor as it had a silent NP-PP branch *man with the*, to which the NP *suitcase* is attached. Finally, the root node of the resulting NP *man with the suitcase* is bound with the NP foot node of the verb as a direct object. Rejecting the direct attachment of *suitcase* to the verb, recovering the NP

²³For simplicity, we leave out binding operations for determiners.

man, inserting the preposition *with* and binding these together requires time, in particular compared to the control condition. Following Hagoort's account of the P600/SPS, the observed effect would precisely reflect the time required by the processor to implement all these steps. An important observation to make is that NP-PP insertion is possible only if the context licenses the appropriate semantic relations between discourse items, in this case only if the information that the man was carrying the suitcase is available from preceding discourse. This entails that when the required semantic relations are not provided by the context, no insertion is possible and thus no P600/SPS is elicited. Instead, if (24c) was processed without the information that the man had a suitcase, the parser would have no alternative binding option than attaching the root NP of *suitcase* with the foot NP direct object node of the verb. In this case, syntactic unification proceeds smoothly, although a violation of semantic restrictions is obtained. This would explain why the P600/SPS was not elicited by (24c) sentences in isolation and why an N400 was obtained instead.

The alternative computational analysis sketched above has several advantages over the account proposed by Nieuwland & van Berkum. First, it is more parsimonious because it treats the observed evoked response within syntax without postulating a semantic illusion effect which, given the absence of behavioral data for the relevant cases, remains undemonstrated. Second, it provides an explanation of the P600/SPS in terms of syntactic binding operations, avoiding revisionary interpretations in terms of monitoring or delayed responses. Third, it shows that in processing (24c) sentences semantic computation is never disrupted and that therefore there is no reason to expect an N400. In particular, animacy restrictions are not violated but actively used to reject the initial attachment of *suitcase* as the direct object of the verb. Discourse semantics is also not violated, since contextual cues are invoked to select the appropriate PP insertion. Fourth, testable predictions can be derived from our analysis: in metonymy or coercion cases like (25a-c) above, a P600/SPS is elicited when context licenses the insertion of the missing syntactic structure, an N400 when no alternative binding option is left to the parser. The predictions that follow from post hoc verbal accounts, such as Nieuwland & van Berkum's, are typically based on analogical inference: since this instance of semantic illusion elicits a P600/SPS, other instances should evoke a similar effect. However, without a computational and algorithmic analysis, the scope of such generalizations is unclear. This shows the benefits of Marr's computational analyses as applied to language, that is more parsimonious theories with increased explanatory and predictive powers. We provide another example below.

3.4.5. *Temporal connectives*. Münte and colleagues [62] recorded event-related potentials while subjects read sentences differing only in the initial temporal connective. Each clause described "a distinct event which is neither logically nor causally related to the other" [62, p. 71], for instance:

- (27) a. After the scientist submitted the paper, the journal changed its policy.
 b. **Before** the scientist submitted the paper, the journal changed its policy.

A true/false comprehension probe followed each critical sentence and was constructed as follows. The sentences were rephrased by changing either the initial temporal connective, its position, the order of the clauses or all three. Subjects were asked to indicate whether the information conveyed by the probe matched with the preceding stimulus sentence. An additional test assessing participants working memory span was performed. Subjects read aloud increasingly larger (2 to 6) sequences of sentences. Immediately after each sequence, participants attempted to recall the last word of each sentence in the order in which these were presented. Scores were used to group subjects (high, medium and low

working memory span). Compared to *after* sentences, *before* sentences elicited a larger sustained negativity, maximal over left anterior sites. At the left frontal electrode (F3) the responses to *before* and *after* sentences diverged as early as 300 ms after sentence onset. The effect lasted for a few seconds and was reliably larger during the second clause. The mean amplitude difference of left frontal negativity between *before* and *after* sentences was positively correlated with memory span: the higher the working memory scores, the more pronounced the negative shift.

According to Münte et al. the observed pattern of effects can be explained along the following lines. As reported above, brain responses to different sentence types diverged within 300 ms following sentence onset, that is as soon as the connective was processed. *After* and *before* access almost immediately different chunks of conceptual knowledge stored in declarative memory: when in sentence-initial position, *after* indicates that the events will be presented in their actual order whereas *before* indicates that the order will be reversed. As to the larger sustained negativity elicited by *before*, it is hypothesized that these require additional discourse-level computations as a consequence of presenting the events in reverse chronological order. Following earlier research, the observed negativity is interpreted as an index of the involvement of working memory in the construction of discourse models.

An important assumption, on which the explanation proposed by Münte and colleagues ultimately relies, is that the order in which events are mentioned in discourse is the only difference between *after* and *before* sentences. This follows from the claim [62, p. 71] that what is mapped onto the meanings of *after* and *before* are temporal relations between events, namely temporal precedence and temporal consequence. However, a review of the linguistic literature on the subject suggests that there is much more to the semantics of temporal connectives than the temporal ordering of events. There is indeed an asymmetry between the inferences licensed by *after* and *before* sentences: *after* always entails that the subordinate clause is true while *before* allows for the subordinate clause to be true or false, depending on the information (contextual or world knowledge) contained in the model. Example (27b) above suggest that *before* sentences, if the subordinate clause is interpreted as false, convey the meaning that the earlier event prevented the later one from occurring. In the situation described by (27b) the scientist might have not been able to submit the paper precisely because of the journal's policy change. A reading in which the subordinate clause is interpreted as false is compatible with the truth of the whole sentence, but the resulting meaning is something quite different from a plain temporal ordering of the events. As we try to argue below, the construction of a discourse model in working memory might be computationally harder for *before* than for *after* sentences.

The analysis is based on the mathematical model of tense and aspect proposed by van Lambalgen & Hamm [90]. In this framework, the tense and aspect of verbs as well as temporal adjuncts introduce so called integrity constraints. These are requirements in the form of obligations and prohibitions that any update of the unification space and in particular any addition to the discourse model based on incoming lexical information should satisfy if a certain condition is fulfilled. *After* sentences must satisfy two integrity constraints. The first is introduced by the main clause and requires a successful update of the unification space such that the event described occurs in the past of the moment of speech, assuming that the tense of the main verb is past tense as in the materials of Münte et al. The second is introduced by the subordinate clause and requires a successful update of the workspace such that the event described occurs before the event mentioned in the main clause:

- $?Happens(E_{main}, t), t < now$ succeeds

- $?Happens(E_{sub}, s), s < t < now$ succeeds

The veridicality of *after* sentences follows from these conditions. *Before* sentences must satisfy three integrity constraints. The first is added to the workspace by the main clause and is identical to the one introduced by main clauses in *after* sentences. The second is introduced by the subordinate clause and can be derived from its *after* counterpart by replacing the obligation with a prohibition. In this way, we account for *after* and *before* as being in some sense ‘converse’ temporal relations. The third requires an update such that the event described by the subordinate clause is located after the event mentioned in the main clause, if no obstacle to its occurrence is described in discourse:

- $?Happens(E_{main}, t), t < now$ succeeds
- $?Happens(E_{sub}, s), s \leq t < now$ fails
- IF E_{main} does not occur in clauses describing obstacles to E_{sub} THEN $?Happens(E_{sub}, t), s < t < now$ succeeds

Notice that, in the interest of symmetry, one could provide an integrity constraint like the third for *after* sentences as well. However, in this case the specification would be at the same time redundant since both events are required to occur by the first two conditions anyway and uninformative because a sequence of events described by an *after* sentence in which the main clause event (i.e. the later one) constitutes an obstacle for the subordinate clause event (i.e. the earlier) violates the principle that causation requires temporal succession. These integrity constraints give the correct predictions for veridicality. Consider again (27b) above. If causal or world knowledge made available in the unification space is such that a journal policy change does not constitute an obstacle for a scientist submitting a paper, then the third condition will force a veridical interpretation of the subordinate clause. If however world knowledge suggests that a policy change is likely to prevent a submission, then forward inference will yield a non-veridical reading, which would be consistent with the second condition above. The competence model thus gives the right predictions as far as linguistic data are concerned.

In order to spell out the predictions of the model at the algorithmic level we would need the full computational machinery detailed by van Lambalgen & Hamm. Informally, the crucial observation is that in the case of *before* sentences, in order to compute the updates satisfying the integrity constraints or, in simpler terms, in order to derive a veridical or a non-veridical reading of the subordinate clause, the unification mechanism has to check whether the information provided by the main clause satisfies the antecedent of the third condition. In *after* sentences the interpretation of the subordinate clause is straightforward because only one option is available, namely the one regimented by the second condition, no matter what will be said by the subsequent main clause. Instead, interpreting a *before* clause can be done in two ways. The first would satisfy the antecedent of the third condition and would lead to a veridical reading of the subordinate clause. The second would exploit forward inference on world knowledge and would yield a non-veridical reading. Which one is eventually computed depends on the information provided by the main clause, which has to be processed before a decision about the subordinate clause can be made. This would explain why *before* sentences are more taxing on the memory resources available for unification, especially during the second clause when the evaluation of both clauses has to be finalized in a coordinated manner.

As it was the case for the semantic illusion example discussed above, our explanation is more parsimonious compared to the account of Münte et al., in which the semantics of temporal connectives is reduced to temporal ordering relations and combined with the hypothesis that reverse order sentences are for some unspecified reason more complex to

process. In our account, an account of ERP data follows from a more appropriate account of the semantic asymmetry between *after* and *before*. Münte et al. might have well missed the point of veridicality because of a failure of intuitions in Hintikka's sense. Indeed, the experimenters assumed a severely restricted view of the contexts in which *before* sentences like (27b) can be uttered. However, in our view the source of the problem lies elsewhere, namely in the attitude toward predictions and explanations in the cognitive neuroscience of language. We are not worried to reiterate that what is missing is a semantic analysis in which the goals of the computation are specified, including not only a representation of the events in their actual order of occurrence, but also an interpretation of *before* clauses as veridical or otherwise. There will be hardly any deep and systematic interaction between linguistics and cognition if post-hoc verbal accounts are not replaced by computational models. There is ultimately a choice to be made between a cognitive science of language as a soft interpretive enterprise or as a set of computational models with testable predictive and explanatory powers. The next section describes a few initial steps in direction of the latter.

4. THE ROAD AHEAD

The starting point of our discussion was a paradox in Chomsky's *Aspects* where, on the one hand, linguistics is regarded as the mentalistic study of competence and, on the other, cognitive data are prevented from contributing in any way to the empirical basis of theories of meaning and grammar. The hypotheses and results on language and cognition considered in sections 1 and 2, on top of other important arguments such as those advanced by cognitive linguists, would suggest to retain mentalism and rethink the competence-performance distinction as well as the use of intuitions as primary evidence in linguistics. As for the latter, we have presented Hintikka's view that introspective judgments, in order to be reliable, should be elicited and controlled in an experiment-like manner. In this way, psychological data become necessary as both a starting point and a partial criterion of adequacy for the theory of competence. Our main point against the competence-performance dichotomy was based on the observation that a wide variety of phenomena is conflated in the notion of performance. This has the effect of unnecessarily complicating the issue of the relationships between competence and performance and hence between linguistics and cognitive (neuro)science. For the same reason, we have rejected Jackendoff's revision of the competence-performance distinction and we have taken Marr's three-level scheme on board. Finally, discussing a few ERP studies on semantic processing, we have tried to emphasize the importance of computational analyses in accounting for experimental data. In the last section of the chapter, we would like to look at one of the emerging issues in contemporary cognitive neuroscience of language, namely the so-called 'binding problem'. This will give us the opportunity to indicate a few directions for future research and to consider what is left out by the approach presented here.

4.1. The binding problem for language. The first formulation of the binding problem is due to von der Malsburg [92], who saw the binding approach to brain function as a response to the difficulties encountered by classical neural networks. Von der Malsburg [93] refers to a well-known example taken from Rosenblatt [74] to illustrate the nature of the issue. Imagine a network for visual recognition constituted by four output neurons. Two fire when a particular shape (either a triangle or a square) is presented and the other two fire depending on the shape's position (top or bottom with respect to a given rectangular display). For instance, if there is a square at the top, the output neuron will read [square, top], if there is a triangle at the bottom, the output will be [triangle, bottom], and so on.

However, if a triangle and a square are presented simultaneously, say the former at the top and the latter at the bottom, the output may read [triangle, square, top, bottom], which is also obtained when the triangle is at the bottom and the square at the top. In other words, the network responds adequately only when single objects are presented in the display. This relates precisely to the binding problem, as von der Malsburg writes:

the neural data structure does not provide for a means of binding the proposition top to the proposition triangle, or bottom to square, if that is the correct description. In a typographical system, this could easily be done by rearranging symbols and adding brackets: [(triangle, top),(square, bottom)]. The problem with the code of classical neural networks is that it provides neither for the equivalent of brackets nor for the rearrangement of symbols. This is a fundamental problem with the classical neural network code: it has no flexible means of constructing higher-level symbols by combining more elementary symbols. The difficulty is that simply coactivating the elementary symbols leads to binding ambiguity when more than one composite symbol is to be expressed. [92, p. 96]²⁴

More famous examples of the binding problem are perhaps bistable figures such as Necker's cube and Jastrow's duck-rabbit, where the exact same visual features of the stimulus lead to two incompatible representations of the object depicted, depending on how these features are bound together. Since the fact that different representations are available depends on the geometrical properties of the figures rather than on the constitution of the perceptual system, as it would be the case for instance for after images [55, pp. 25-26], bistability requires an explanation at the computational level of analysis. These examples suggest that binding in the visual domain is a genuine information processing problem, but what about language?

Although binding mechanisms are likely to differ across cognitive domains, there exist analogous cases of bistability in language processing. Consider the following sentences:

- (28) a. The woman saw the man with the binoculars.
b. Respect remains.

Example (28a) has two possible syntactic representations, one in which the phrase *with the binoculars* is a PP attached to the NP *the man* (i.e. the man had binoculars), another in which it modifies the VP (i.e. the woman used binoculars to see the man). Again, the same features of the stimulus can lead to two entirely different representations, depending on which attachment option is eventually chosen. These sentences typically result in larger P600/SPS components, suggesting that syntactic binding is indeed a problem for the brain. Example (28b) has two possible syntactic *and* semantic representations, one as a request where *respect* is the verb and *remains* the object noun, the other as a statement where *respect* is the object noun and *remains* the verb.²⁵ There may be some superficial similarities between visual and linguistic bistability, such as the fact that in both cases we seem to 'flip'

²⁴We cannot enter the details of the various attempted solutions to Rosenblatt's problem. The reader can refer to von der Malsburg [93] for a proposal in line with the binding hypothesis and to Riesenhuber & Poggio [72] for an alternative approach.

²⁵Example (28a) bears some resemblance to the case of *The sheriff saw the indian and the cowboy ...* considered above. However, the latter is a slightly different (and possibly more involved) case of syntactic bistability. In fact, what is bistable there is a *partial* syntactic representation and the ambiguity is resolved when the upcoming phrase *noticed the horses in the bushes* is processed. A parallel case for example (28b) would be *Before the scientist submitted the paper ...*, which can be seen as temporarily bistable while the subsequent main clause typically (although not always) disambiguates the semantic value of the subordinate.

between the two incompatible representations, but the deep similarity, in our view, is that bistability has to be defined at the topmost level of analysis, as Marr himself pointed out [55]. To extend Marr's line of thought, we suggest that the binding problem for language is best formulated at the computational level, although attempted solutions are bound to involve dramatic changes at all levels of analysis, including neural implementation [28, 29]. As we have seen in section 3.4.4, a clear formulation and solution of binding as a computational problem has been proposed for the case of syntax [27]. However, a parallel piece of work for semantics is still missing. Our next task is to motivate and outline an account of semantic unification.

4.2. The semantic binding problem: toward a model of semantic unification. The trenchant separation between formal semantics and the psychological study of meaning has persisted from the early days of Frege until relatively recently. For that and other more technical reasons, the integration of semantic theory and cognitive science is perhaps too ambitious a goal for present-day research. In fact, there are obstacles on both sides of the barricade, with semanticists proposing theories based on not necessarily computable functions²⁶ and cognitively implausible assumptions and psycholinguists adopting naive and computationally unclear notions such as selection, integration, and the like. Claiming that such obstacles can be easily identified and removed would be downplaying the difficulty of the task. Nevertheless, we believe it is possible to list some of the main ingredients of a model of semantic binding without giving a definite recipe for cooking them up. Below we focus on two such ingredients, namely a cognitively realistic notion of semantic model and a view of reasoning as one key binding mechanism at the sentence and discourse levels.²⁷

Traditional semantic theory makes a distinction between the meaning of an expression, the structures in which the expression is interpreted and the representations associated with the interpretation process. According to the standard Fregean view [21], for instance, the different ideas that a painter, a horseman and a zoologist connect with the name *Bucephalus* are not to be confused with the sense of the word (the thought it expresses) or with its reference (Alexander the Great's horse). The sense and reference of an expression are objective entities, belonging to the domain of thoughts and the real world respectively, while ideas or mental representations are subjective, fluctuating and generally recalcitrant toward any attempt of description or comparison. Crucially, in formal semantics expressions are interpreted directly into real-world structures and mental representations are seen as little more than a by-product of the interpretation process. The situation is different in more recent accounts. In cognitive semantics, for instance, the relation between an expression and the real world is broken down into two: a properly semantic relation between an expression and a mental representation and a causal relation between the representation and the real world. In Discourse Representation Theory (DRT), interpretation proceeds via the construction of a cognitive model of discourse intermediate between the syntactic representation and the world. It has been observed that semantic representations - roughly in the sense of cognitive linguistics and DRT - have to have particular properties in order to fit with a computationally viable notion of interpretation [34, 90, 81]. To illustrate this point, consider the following sentences:

²⁶See for instance the treatment of Frege's notion of sense in possible worlds semantics.

²⁷A third important ingredient is a more appropriate (viz. computational) account of the Fregean notions of sense and reference. We cannot enter the details of the issue here. However, we refer the reader to Moschovakis' [61] original proposal to identify the sense of an expression with the algorithm that computes the expression's denotation. Extensions and applications of this idea can be found in [89, 90].

- (29) a. John was crossing the street when he was hit by a truck.
 b. John was crossing the street when he saw his friend Paul.

In (29a) we can presume that John did not reach the other side of the street because the accident prevented him to do so. However, in (29b) we have no positive information on whether seeing Paul had a terminating effect on the crossing activity. The problem is precisely how the cognitive system and ultimately the brain handle the absence of positive information. It seems improbable that statements specifying all the possible effects and non-effects of events are stored in declarative memory, since this would lead to a combinatorial explosion which would make semantic binding extremely hard, if not impossible. What seems more plausible, at least from a computational point of view, is that the brain treats information which is not provided by (or cannot be derived from) discourse or world knowledge as false, until evidence for the contrary is provided. This principle, known in non-monotonic logic as closed world reasoning, leads to discourse models which are *minimal* in the sense that what is represented is restricted to positive information provided by discourse and world knowledge as well as their logical consequences. Minimal models seem particularly appealing in the framework of the binding problem because (i) in all relevant cases, for instance sentences (29a-b) above, there exist a unique minimal model interpreting discourse and (ii) the procedure for obtaining minimal models from the sense of the expressions involved is fully computable. The uniqueness and computability of minimal models is consistent with the limited storage capacity of working memory, which we assume is the ‘unification space’ for semantics. We have very roughly demarcated the type of structures that could be the result of semantic binding operations. It is now time to have a closer look at these operations themselves.

Consider examples (29a-b). Let us assume that interpretation proceeds incrementally as words are encountered. The brain has to unify information provided by individual words, which includes not only lexico-semantic specifications or core meanings but also temporal and aspectual information carried by verb phrases and represented in the form of integrity constraints, causal knowledge stored in memory and made salient by particular phrases and assumptions concerning negative evidence for the occurrence of certain classes of events such as obstacles to goals. Processing the progressive clause *John was crossing the street* leads the introduction of an integrity constraint (see 3.4.5) stating that every update of the unification space should be consistent with the information that John’s crossing activity was taking place at some point in time in the past of the moment of speech (see [90, pp. 156-160] for details). Up to this stage there is no positive evidence that John encountered any obstacle while crossing the street. Therefore, closed world reasoning would lead to a minimal model in which, after an unspecified amount of time, the crossing activity brings John to the other side of the street (see [90, pp. 159-161] for a rigorous statement and a computational proof). The minimal model is extended when the subordinate clause in (29b) is processed since, again, no positive evidence of obstacles is provided and the inference that John eventually reached the other side of the street can still be drawn. Notice that, without closed world reasoning, semantic binding would soon reach a dead end: if the system cannot assume that seeing Paul had no effect on the crossing activity, then nothing follows about John having reached the other side of the street, which is not what is implied by these sentences.²⁸ An important consequence of the theory is that minimal models can

²⁸On this seemingly problematic point, we have conducted a preliminary behavioral study asking subjects (24 Dutch native speakers, mean age 26.8, age range 22-41) whether by reading sentences such as (29a-b) above they could assume that the relevant goal state (say, John having reached the other side of the street) was attained. Progressive clauses presented in isolation (that is, without the subsequent *when* clause) were judged as entailing

be recomputed on-line when previously missing positive information is provided by discourse, as is the case in (1a). On the assumption that the occurrence of the goal state (John having reached the other side of the street) is inferred when the progressive clause is processed,²⁹ the information that John was hit by a truck forces the system to withdraw the inference that he had reached the other side of the street. Inference on minimal models is thus defeasible, that is it can be cancelled if evidence for the contrary is provided.

5. CONCLUSION

The purpose of this chapter was to examine the relations between language and cognition from a broad methodological standpoint. Given our initial choice, we have rapidly shifted the focus of the discussion toward the relations between linguistic theory and cognitive science. We have seen that linguists belonging to different schools tend to rely exclusively on introspective judgments as the evidential basis of theories of meaning and grammar. However, intuitions are not generally reliable and are only loosely aligned with the formal notions of synonymy and well-formedness. Therefore, experimental data seem at least a necessary complement of introspective judgments. Analogously, experimenters studying language comprehension and production tend to base their accounts of behavioral and brain data more on pre-theoretic intuitions of the properties of the relevant linguistic strings than on linguistically motivated processing models. We have argued that linguistics can play a role in psycholinguistics and cognitive neuroscience if it is regarded as the core part of a computational theory in Marr's sense. Theories of phonology, syntax and semantics can provide descriptions of the data structures that are computed by the human brain. In other words, linguistics would amount to an analysis of the goals and subgoals of the relevant comprehension and production tasks. In general, as others have argued before us, a more intense exchange at all levels between linguistics and cognitive science is required if we are to understand human communication. Crucially, interactions between disciplines should take place within a shared framework, accommodating the individual contributions in the larger picture. We have proposed Marr's classic three level scheme as one such framework.

As to the binding problem for language, we have presented and applied a computational model of syntactic unification and we have sketched an account of semantic binding based on minimal models and defeasible reasoning. This is a first but nevertheless important step, for three reasons. First, the proposed account is computationally explicit. We have argued that the binding problem is best formulated at Marr's topmost level of analysis. So what is needed is a theory in which the data structures computed by the human brain are characterized in a rigorous manner, motivated by cognitive considerations and supported by successful analyses of linguistic data. For instance, an issue which a model of semantic binding should address is how linguistic meanings and world knowledge are combined to arrive at a coherent interpretation of discourse. As we have hinted at above discussing non-veridical *before* and the progressive, minimal models and closed world reasoning seem up to this task. Second, the model is cognitively plausible. Minimal models seem to be the appropriate choice given the limited working memory resources. In particular, computing with unique minimal models seems more efficient than handling largely underspecified or even several models simultaneously. We have also assumed that defeasible reasoning

the occurrence of the goal state in 79.58% of the cases, sentences like (29b) in 72.92% of the cases and sentences like (29a) in 22.86% of the cases. These results suggest that subjects take progressive sentences, when presented in isolation or when no positive evidence of terminating events is given, as entailing the culmination of the relevant action or event. This is consistent with the minimal models account of semantic binding we are sketching here.

²⁹This depends on the actual algorithms and memory mechanisms subserving the computation of minimal models. We are currently conducting an EEG/ERP study to address this issue.

was initially developed for planning needs. Efficient planning requires that unavailable information is treated as false and that action sequences toward a goal can be readjusted when obstacles are encountered along the way. Discourse processing might have exploited a computational machinery which was already available [78, 90, 81], although some restructuring is not unlikely. Third, the model provides testable predictions for behavioral and electrophysiological data, although dedicated experimental work is still lacking. A viable strategy, in our view, is to spell out a theory of semantic binding at the three levels of Marr's scheme, trying to accommodate experimental data and computational theories of lexical, sentence and discourse meaning. We grant that a lot more than the what we have assumed here must be taken on board. However, minimal models and defeasible reasoning seem to us essential for the purposes of the enterprise.

REFERENCES

- [1] R.A. Berman and D.I. Slobin (eds.). *Relating events in narrative*. Lawrence Erlbaum, Hillsdale, N.J., 1994.
- [2] D. Bolinger. Judgments of grammaticality. *Lingua*, 21:34–40, 1968.
- [3] G. Boole. *An Investigation of the Laws of Thought*. Dover, 1958.
- [4] J.D. Bransford, J.R. Banks, and J.J. Franks. Sentence memory: a constructive versus an interpretive approach. *Cognitive Psychology*, 3:193–209, 1972.
- [5] T. Burge. *Truth, Thought, Reason, Essays on Frege*. Clarendon Press, 2005.
- [6] M.N. Carminati, L. Frazier, and K. Rayner. Bound variables and c-command. *Journal of Semantics*, 19:1–34, 2000.
- [7] P. Carruthers. *Language, Thought and Consciousness*. Cambridge University Press, 1996.
- [8] P. Carruthers. The cognitive functions of language. *Behavioral and Brain Sciences*, 25:657–725, 2002.
- [9] N. Chomsky. *The Logical Structure of Linguistic Theory*. MIT Press, 1955.
- [10] N. Chomsky. *Syntactic Structures*. De Gruyter-Mouton, 1957.
- [11] N. Chomsky. *Aspects of the Theory of Syntax*. MIT Press, 1965.
- [12] N. Chomsky. *Lectures on Government and Binding*. Foris, 1981.
- [13] N. Chomsky. *The Minimalist Program*. MIT Press, 1995.
- [14] N. Chomsky, A. Belletti, and L. Rizzi. *On Nature and Language*. Cambridge University Press, 2002.
- [15] W. Croft and D.A. Cruse. *Cognitive Linguistics*. Cambridge University Press, 2004.
- [16] D. Dennett. *Consciousness Explained*. Penguin, 1991.
- [17] E. Donchin. Event-related brain potentials: A tool in the study of human information processing. In *Evoked Brain Potentials and Behavior*, pages 13–88. Plenum Press, 1979.
- [18] G. Fauconnier. Methods and generalizations. In *Cognitive Linguistics: Foundations, Scope, and Methodology*, pages 95–128. Mouton-De Gruyter, 2000.
- [19] S. Featherston, M. Gross, T.F. Münte, and H. Clahsen. Brain potentials in the processing of complex sentences: An erp study of control and raising constructions. *Journal of Psycholinguistic Research*, 29(2):141–154, 2000.
- [20] J. Fodor. *The Modularity of Mind*. MIT Press, 1983.
- [21] G. Frege. On sense and reference. In M. Black P. Geach, editor, *Translation from the Philosophical Writings of Gottlob Frege*. Blackwell, 1980.
- [22] H. Geurts. *Executive functioning profiles in ADHD and HFA*. PhD thesis, Free University (Amsterdam), 2003.
- [23] A. Goldberg. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, 1995.
- [24] A. Goldberg. *Constructions at Work: the nature of generalization in language*. Oxford University Press, 2006.
- [25] P.M. Greenfield. Language, tools and the brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and brain sciences*, 14:531–595, 1991.
- [26] P. Hagoort. The shadows of lexical meaning in patients with semantic impairments. In B. Stemmer and H.A. Whitaker, editors, *Handbook of Neurolinguistics*, pages 235–248. Academic Press, 1998.
- [27] P. Hagoort. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, 20:S18–29, 2003.
- [28] P. Hagoort. On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9:416–423, 2005.

- [29] P. Hagoort. The binding problem for language and its consequences for the neurocognition of comprehension. In N. Pearlmutter E. Gibson, editor, *The Processing and Acquisition of Reference*. MIT Press, 2006.
- [30] P. Hagoort and C. Brown. Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, and K. Rayner, editors, *Perspectives on Sentence Processing*, pages 45–81. Lawrence Erlbaum Associates, 1994.
- [31] P. Hagoort, C. Brown, and L. Osterhout. The neurocognition of syntactic processing. In *The Cognitive Neuroscience of Language*, pages 273–307. Oxford University Press, 2001.
- [32] P. Hagoort, C.M. Brown, and J. Groothusen. The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, 8:439–483, 1993.
- [33] P. Hagoort, L. Hald, M. Bastiaansen, and K.M. Petersson. Integration of word meaning and world knowledge in language comprehension. *Science*, 304:438–41, 2004.
- [34] F. Hamm, H. Kamp, and M. van Lambalgen. There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, ?(?):??–??, 2006.
- [35] J. Hintikka. Intuitions and philosophical method. *Revue Internationale de Philosophie*, 35:74–90, 1981.
- [36] J. Hintikka. The emperor’s new intuitions. *Journal of Philosophy*, 96:127–147, 1999.
- [37] J. R. Hurford. The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, 26:261–316, 2003.
- [38] R. Jackendoff. On beyond zebra: the relation of linguistic and visual information. *Cognition*, 26:89–114, 1987.
- [39] R. Jackendoff. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, 2003.
- [40] R. Jackendoff. Precis of foundations of language: brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, 26:651–65; discussion 666–707, 2003.
- [41] P.N. Johnson-Laird. Mental models in cognitive science. *Cognitive Science*, 4:71–115, 1980.
- [42] T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T.J. Sejnowski. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14:166–185, 2001.
- [43] A. Kim and L. Osterhout. The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52:205–225, 2005.
- [44] W. Kohler. *The mentality of apes*. Harcourt Brace and World, New York, 1925.
- [45] G.R. Kuperberg, T. Sitnikova, D. Caplan, and P. J. Holcomb. Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17:117–129, 2003.
- [46] M. Kutas and S.A. Hillyard. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207:203–205, 1980.
- [47] M. Kutas and S.A. Hillyard. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163, 1984.
- [48] W. Labov. When intuitions fail. *Papers from the Parasession on Theory and Data in Linguistics*. Chicago Linguistic Society, 32:77–106, 1996.
- [49] W. Levelt. Some psychological aspects of linguistic data. *Linguistische Berichte*, 17:18–30, 1972.
- [50] W. Levelt, J. van Gent, A. Haans, and A. Meijers. Grammaticality, paraphrase, and imagery. In S. Greenbaum, editor, *Acceptability in Language*, pages 87–101. Mouton, 1977.
- [51] D. Lewis. General semantics. *Synthese*, 22:18–67, 1970.
- [52] S. Luck. *An Introduction to the Event-Related Potential Techniqu*. MIT Press, 2005.
- [53] A.R. Luria. *Cognitive Development: its social and cultural foundations*. Harvard University Press, Cambridge, MA, 1976.
- [54] A. Marantz. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22:429–445, 2005.
- [55] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, 1982.
- [56] D. Marr and T. Poggio. From understanding computation to understanding neural circuitry. *MIT AI Memos*, 1976.
- [57] J.J. McCarthy and A. Prince. The emergence of the unmarked: Optimality in prosodic morphology. In *Proceedings of NELS*, volume 24, pages 333–379. GLSA, University of Massachusetts, Amherst, 1994.
- [58] B. McGonigle, M. Chalmers, and A. Dickinson. Concurrent disjoint and reciprocal classification by *cebus apella* in serial ordering tasks: evidence for hierarchical organization. *Animal Cognition*, 6(3):185–197, 2003.
- [59] R. McKinnon and L. Osterhout. Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. *Language and Cognitive Processes*, 11(5):495–523, 1996.

- [60] C.T. McMillan, R. Clark, P. Moore, C. Devita, and M. Grossman. Neural basis of generalized quantifier comprehension. *Neuropsychologia*, 43:1729–1737, 2005.
- [61] Y. Moschovakis. Sense and denotation as algorithm and value. In J. Oikkonen and J. Vaananen, editors, *Lecture Notes in Logic*, 2. Springer, 1994.
- [62] T.F. Münte, K. Schiltz, and M. Kutas. When temporal terms belie conceptual order. *Nature*, 395:71–3, 1998.
- [63] M.S. Nieuwland and J.J.A. Van Berkum. Testing the limits of the semantic illusion phenomenon: Erps reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24:691–701, 2005.
- [64] L. Osterhout and P. Holcomb. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31:785–806, 1992.
- [65] C. Phillips, N. Kazanina, and S.H. Abada. Erp effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22:407–428, 2005.
- [66] F. Pulvermuller. Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6:576–582, 2005.
- [67] F. Pulvermuller, M. Harle, and F. Hummel. Walking or talking? behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, 78:143–168, 2001.
- [68] F. Pulvermuller, O. Hauk, V.V. Nikulin, and R.J. Ilmoniemi. Functional links between motor and language systems. *European Journal of Neuroscience*, 21:793–797, 2005.
- [69] K.L. Purvis and R. Tannock. Language abilities in children with attention deficit disorder, reading disabilities, and normal controls. *Journal of Abnormal Child Psychology*, 25(2):133–144, 1997.
- [70] R. Quiñan Quiroga and H. García. Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 114:376–390, 2003.
- [71] W. V. Quine. Methodological reflections on current linguistic theory. *Synthese*, 21:386–398, 1970.
- [72] M. Riesenhuber and T. Poggio. Are cortical models really bound by the ‘binding problem’? *Neuron*, 24:87–93, 1999.
- [73] M. D. Root. Speaker intuitions. *Philosophical Studies*, 29:221–234, 1976.
- [74] F. Rosenblatt. *Principles of Neurodynamics: Perceptions and the Theory of Brain Mechanisms*. Spartan Books, 1961.
- [75] S. Scribner. *Mind and social practice*. Cambridge University Press, Cambridge, 1997.
- [76] M. Singer. Discourse inference processes. In M.A. Gernsbacher, editor, *Handbook of Psycholinguistics*. Academic Press, 1994.
- [77] M.J. Spivey and Gonzalez-Marquez M. Rescuing generative linguistics: Too little, too late? *Behavioral and Brain Sciences*, 26:690–691, 2003.
- [78] M. Steedman. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 25:723–753, 2002.
- [79] M. Steedman. Plans, affordances and combinatory grammar. *Linguistics and Philosophy*, 25(5–6):725–753, 2002.
- [80] K. Stenning. How did we get here? a question about human cognitive evolution. Frijda Lecture, University of Amsterdam, 2003.
- [81] K. Stenning and M. van Lambalgen. Semantic interpretations as computation in non-monotonic logic: the real meaning of the suppression task. *Cognitive Science*, 2005.
- [82] K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science*. MIT Press, 2006.
- [83] K. Stenning and M. van Lambalgen. *Human reasoning and cognitive science*. MIT University Press, Cambridge, MA., 2006.
- [84] T. Trabasso and N.L. Stein. Using goal-plan knowledge to merge the past with the present and future in narrating events on line. In M.H. Haith, J.B. Benson, R.J. Roberts, and B.F. Pennington, editors, *The development of future-oriented processes*, pages 323–352. University of Chicago Press, 1994.
- [85] J.J. van Berkum, C. Brown, P. Hagoort, and P. Zwitserlood. Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, 40:235–248, 2003.
- [86] J.J. van Berkum, C.M. Brown, and P. Hagoort. Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41:147–182, 1999.
- [87] J.J. van Berkum, P. Hagoort, and C.M. Brown. Semantic integration in sentences and discourse: evidence from the n400. *Journal of Cognitive Neuroscience*, 11:657–671, 1999.
- [88] J.J.A. van Berkum, P. Zwitserlood, P. Hagoort, and C.M. Brown. When and how do listeners relate a sentence to the wider discourse? evidence from the n400 effect. *Cognitive Brain Research*, 17:701–718, 2003.
- [89] M. van Lambalgen and F. Hamm. Moschovakis’ notion of meaning as applied to linguistics. In S. Friedman M. Baaz and J. Krajček, editors, *Logic Colloquium ’01, ASL Lecture Notes in Logic*. A.K. Peters Publishers, 2004.

- [90] M. van Lambalgen and F. Hamm. *The Proper Treatment of Events*. Blackwell, 2004.
- [91] M. van Lambalgen and F. Hamm. *The proper treatment of events*. Blackwell, Oxford and Boston, 2004.
- [92] C. von der Malsburg. The correlation theory of brain function, 1981.
- [93] C. von der Malsburg. The what and why of binding: the modeler's perspective. *Neuron*, 24:95–104, 1999.
- [94] T. Vosse and G. Kempen. Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75:105–143, 2000.
- [95] D. Wilkinson and P. Halligan. The relevance of behavioral measures for functional-imaging studies of cognition. *Nature Neuroscience*, 5:67–73, 2004.
- [96] R. Willems, A. Özyürek, and P. Hagoort. When language meets action: the neural integration of gesture and speech. Under review, 2005.

Processing temporal constraints: An ERP study

Giosuè Baggio

Author affiliation:

Giosuè Baggio

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen

Address correspondence to:

Giosuè Baggio

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen

P.O. Box 9101

NL-6500 HB Nijmegen

The Netherlands

Phone: +31.024.3610980

Fax: +31.024.3610989

E-mail: g.baggio@fcdonders.ru.nl

Abstract

This paper investigates how linguistic expressions of time, in particular temporal adverbs and verb tense morphemes, are used to establish temporal reference at the level of brain physiology. First, a formal semantic analysis of tense and temporal adverbs is outlined. It is argued that computing temporal reference amounts to solving a constraint satisfaction problem. Next, the results of an event-related potential (ERP) study are presented which suggest that violations of verb tense ('Last Sunday Vincent *paints* the window frames of his country house.') evoke larger left-anterior negativities (LANs) compared to sentences in which verb tense is correct. The semantic analysis is finally combined with a computational, unification-based model of parsing to provide a functional account of the ERP data.

Introduction

Reference to time is ubiquitous in natural languages, to the point that nearly every assertion involves the location of some event within a temporal coordinate system. Much attention has been paid to the expressions used to encode temporal information, which range from prepositional phrases (e.g. *before the dawn*) to verb suffixes (e.g. ‘-ed’ in the English regular simple past), to measure phrases borrowed from mathematics (e.g. ‘ 10^{-35} seconds into the expansion phase’). Linguists have provided detailed accounts of the functioning of these devices, often using the formal tools made available by modern logic. However, comparatively little is known about how expressions of time are represented and used to compute temporal reference at the level of brain physiology. My aim here is to contribute to this line of research by combining a formal analysis of temporal expressions with ERP data on tense processing.

A semantic analysis of tense and temporal adverbs

In this section I shall be concerned with a formal semantic analysis of sentences in which tense is correctly or incorrectly used, such as:

- (1) a. Last Sunday Vincent painted the window frames of his country house.
- b. *Last Sunday Vincent paints the window frames of his country house.

These sentences recognizably form a minimal pair of the kind customarily used in language processing experiments, and the formal treatment of tense outlined below is precisely intended to drive the interpretation of the ERP data reported later on. The main theoretical claim will be that processing sentences such as (1a) involves solving a constraint satisfaction problem in which the relevant temporal constraints are introduced by the semantics of the adverb *last Sunday* and of the verb *painted*. Conversely, processing a sentence such as (1b) is claimed to lead to a failure to simultaneously solve such constraints. Although the analysis developed here brings semantics center stage, tense has ramifications in morphology which can hardly be ignored. Before I turn to semantics, some notes on morphology are therefore in order.

Morphology

Comrie (1985) defined verb tense as the “grammaticalization of location in time”. Unpacked, this definition reads that tense is concerned with the location of events in time – or, possibly, a cognitive representation of time – and that the expression of time is (i) obligatory and (ii) morphologically bound on verbs. However, this can only be taken to imply that verbs obligatorily carry temporal information via morphemes, though not that there are morphological rules, which determine, for any given verb form, whether tense is correctly used. Let me try to clarify this point with reference to examples (1).

The verb forms *painted* in (1a) and *paints* in (1b) convey, via the suffixes ‘-ed’ and ‘-s’, the information that the painting event is located, respectively, in the past and in

the present of the moment of speech. Importantly, both verb forms are morphologically correct, in the sense in which, for instance, the over-regularized verb form *goed* (past of ‘to go’) is not. What makes *paints* incorrect in the context of (1b) is a mismatch between its semantics, that is, the fact that it refers (at least in this context) to a present event, and that of the adverb *last Sunday*, which defines a temporal window in the past of the moment of speech in which the main event is to be located.

From these considerations, two views on the processing consequences of the violation in (1b) ensue. Under the first, (1b) is perceived by the language system as a semantic violation, and not as a morphological one, because, as we have seen, *paints* is in fact morphologically well-formed. Under the second, *paints* in (1b) is perceived as a morphological anomaly, even though its origin is semantic. This option entails the additional hypothesis that the system is endowed with an interface component (Jackendoff, 2002) mediating between semantics and morphology which, on the basis of the meaning of *last Sunday*, constrains the set of suffixes which the verb stem *paint-* – or any other verb stem for that matter – can take. In this paper I wish to remain neutral with respect to this issue, despite its obvious importance. Instead, I shall consider the processing consequences of the observation that, regardless of whether (1b) is perceived as containing a semantic or a morphological anomaly, it is the mismatch between the *semantic* features of the adverb and the verb which determines that tense is used incorrectly.

Semantics

In the coming paragraphs I will outline an analysis of sentences like (1a) and (1b), trying to add some formal detail to the claim that tense violations are semantic in nature. The analysis is based on a lightweight adaptation of the mathematical theory of tense and aspect of van Lambalgen and Hamm (2004). The formalism they present and apply is an Event Calculus, reformulated using Constraint Logic Programming, which has its origins in the design of agents capable of planning in robotics. The result is something quite different from the event calculi of the Davidsonian tradition, such as those used by Parsons (1990) and Larson and Segal (1995).

The Event Calculus of van Lambalgen & Hamm is a many-sorted predicate logic with sorts for eventualities viewed either perfectly, called *events*, or imperfectly, called *fluents*. As we shall see soon, fluents can be used to represent processes and temporal intervals, such as the period denoted by *last Sunday*. Temporal and causal relations are formalized using *predicates* and *constraints*, combined in formulae called either *clauses* or *integrity constraints*. Here I will use two predicates only: *Happens*(e, t), which means that event type e has a token at time t , and *HoldsAt*(f, t), meaning that the fluent f holds at time t . Constraints are either equalities ($x=y$) or inequalities ($x\leq y$) relating the values of the temporal variables that occur as arguments of the predicates.

Discourse comprehension involves the construction of a model making discourse true (van Lambalgen & Hamm, 2004; Stenning & van Lambalgen, 2005). As each

word is processed, an *estimate* of the meaning intended by the speaker is computed, that is, a representation making what is heard or read true and consistent with the grammar and world knowledge. The process is incremental, in that each new incoming word may result in an update of the model computed during the preceding stage. While lexical meanings are formalized as sets of clauses, of which an example will be given soon, semantic updates are regimented by integrity constraints. These are statements that can take the form of either obligations ‘? Φ succeeds’, forcing an update of the model satisfying Φ , or prohibitions ‘? Φ fails’, blocking updates of the model satisfying Φ .

The tenses can be treated as integrity constraints instructing the system to update the current discourse model so as to locate the relevant event in the past, present or future of the moment of speech (van Lambalgen & Hamm, 2004). The following integrity constraints may serve as a first approximation of the meaning of the past tense:

(2) a. ? *Happens*(e, t) $\wedge t < now$ succeeds

b. ? *Happens*(e, t) $\wedge t = now$ fails

While (2a) forces the system to update the current discourse model such that e is located at t , in the past of the moment of speech *now*, (2b) captures the implicature that the event e lies entirely in the past. However, as has often been observed (Partee, 1973), the past tense is anaphoric, for it requires an anchor expression, such as a temporal adjunct, in the preceding discourse context. (Forward-looking temporal

reference seems possible too, but let us ignore this possibility here.) The sentence *Vincent painted the window frames of his country house*, uttered without a prior temporal context, would seem under-informative (Grice, 1989). This suggests that (2) should be refined as follows:

(3) a. $? HoldsAt(f, t) \wedge Happens(e, t) \wedge t < now$ succeeds

b. $? HoldsAt(f, t) \wedge Happens(e, t) \wedge t = now$ fails

(4) $? HoldsAt(f, t) \wedge Happens(e, t) \wedge t = now$ succeeds

The meaning of the present tense is captured by the following integrity constraint (see van Lambalgen & Hamm (2004) for a more careful analysis):

The fluent f in (3) and (4) has to be unified with material from the preceding discourse context, for instance a temporal adverb. *Last Sunday* can be analyzed as follows (Hamm, Kamp, & van Lambalgen, 2006). To begin with, the time line is partitioned in seven segments of equal length, corresponding to the days of the week: fSu, fMo, \dots, fSa . Next, the meaning of the adverb is given as in (5), where $fCPSu$ denotes the closest past Sunday:

(5) $HoldsAt(fSu, s) \wedge s < now \wedge |now-s| \leq 7 \text{ days} \rightarrow HoldsAt(fCPSu, s)$

Unification works as follows. Upon encountering a sentence such as (1a) or (1b), the system is incrementally confronted with a constraint satisfaction problem.

Consider first (1a). Processing the adverb *last Sunday* results in a model in which the query $?HoldsAt(fCPSu, t)$ succeeds, that is, in which the antecedent of (5) holds.

When the verb *painted* is processed, the system tries to satisfy (3). This is done by unifying f in (3) with $fCPSu$ in (5) ($f = fCPSu$) and t in (3) with s in (5) ($s = t$). All constraints introduced by (1a) can be satisfied, and no inconsistency follows. Let us now turn to (1b). Processing the adverb *last Sunday* leads, once again, to a model in which the antecedent of (5) holds. When the verb *paints* is processed, the system tries to satisfy (4). However, unifying s with t ($s = t$) now leads to an inconsistency: while $s < now$ has been satisfied when processing the adverb, (4) requires on the contrary that $t = now$ (or equivalently, under the proposed unification, $s = now$) succeeds; the two constraints cannot be satisfied simultaneously. This suggests a precise definition of ‘tense violation’: there is a mismatch between an adverb and verb tense, or two verb tenses, when the unification of the temporal variables set up by the relevant expressions fails.

Consequences for ERP studies of tense

The definition just proposed can be used in a critical assessment of the ERP (event-related potential) literature on tense. The first attempt to investigate tense violations using ERPs was made by Kutas and Hillyard (1983). In this experiment, a group of English speakers was presented with sentences like the following:

(6) a. Most of the earth’s weather happens in the bottom layer of the atmosphere *called* the troposphere.

b. *Most of the earth's weather happens in the bottom layer of the atmosphere *calls* the troposphere.

c. The eggs and meat of this turtle are *considered* choice food by many people.

d. *The eggs and meat of this turtle are *consider* choice food by many people.

e. This allows them to *stay* under water for a longer period.

f. *This allows them to *stayed* under water for a longer period.

Incorrect sentences elicited a positive shift peaking at about 300 ms following the verb and approaching significance at parietal sites, and a negative wave in the 300-400 ms window, suggestive of an N400 effect (Kutas & Hillyard, 1980). Regardless of the functional meaning of these effects, there appear to be two problems with this study. First, (6b), (6d) and (6f) do not instantiate tense violations, as none of these sentences contains a temporal expression, such as a temporal adverb or a temporal preposition phrase, with which verb tense fails to agree. Second, anomalies are realized in quite different ways. In (6b) the present indicative *calls* replaces the past participle *called*, in (6d) the infinitive *consider* replaces the past participle *considered*, and in (6e) the past participle *stayed* replaces the infinitive *stay*.

Osterhout and Nicol (1999) investigated the ERP correlates of “verb tense violations” in modal constructions, presenting to a group of English speakers sentences like:

(7) a. The cats won't *eat* the food that Mary leaves them.

- b. *The cats won't *eating* the food that Mary leaves them.
- c. The expensive ointment will *cure* all known forms of skin disease.
- d. *The expensive ointment will *curing* all known forms of skin disease.
- e. The new fighter planes can *fly* faster than anyone had expected.
- f. *The new fighter planes can *flying* faster than anyone had expected.

In the 300-500 ms window, incorrect sentences were more positive at posterior sites of the mid-line, whereas the LAN (left anterior negativity) elicited by anomalous verbs did not differ significantly from the effect evoked by correct verbs. In the 500-800 ms time interval, incorrect sentences were more positive at mid-line and posterior sites, consistent with the distribution of the P600 (Osterhout & Holcomb 1992; Hagoort, Brown, & Groothusen, 1993). ERPs elicited by sentence-final words in the violation condition were more negative-going compared to those observed in correct sentences, they started 200 ms following the sentence-final word and continued throughout the epoch. Also in this case, it should be noted that sentences (7b, d, f) do not contain tense violations: all main verbs lack an anchor point in the preceding discourse with respect to which tense agreement can be evaluated.

In an ERP study by Allen, Badecker, and Osterhout (2003), “syntactic (tense) violations” were investigated in sentences containing either high-frequency (HF) or low-frequency (LF) verbs:

- (8) a. The man will *work* on the platform. (HF)

- b. *The man will *worked* on the platform. (HF)
- c. The man will *sway* on the platform. (LF)
- d. *The man will *swayed* on the platform. (LF)

Correct low-frequency verbs elicited a larger N400, ungrammatical verbs a larger P600, and low-frequency anomalous verbs a bi-phasic increase of N400 and P600. In the 500-900 ms window a significant main effect for grammaticality was found, maximal on posterior sites as is typical of the P600. Again, (8b, d) do not contain tense violations, contrary to the authors' claim that the "sentences 'He will walked' and 'He will swayed' are equally and unconditionally ill-formed with respect to tense" (Allen et al., 2003).

In summary, the studies of Kutas and Hillyard (1983), Osterhout and Nicol (1999) and Allen et al. (2003) did not bring into play genuine tense violations. In none of these experiments were stimulus sentences such that the anomalous verb located the event described by the main clause outside a frame specified by an anchoring temporal expression.

Genuine tense violations have been investigated in at least two ERP studies. Fonteneau, Frauenfelder, and Rizzi (1998) presented a group of French speakers with sentences like:

- (9) a. Demain l'étudiant *lira* le livre.
'Tomorrow the student will read the book.'

b. *Demain l'étudiant *lisait* le livre.

'Tomorrow the student read the book.'

The adverb *demain* 'tomorrow' specifies a temporal frame within which the eventuality denoted by the main clause is taken to occur. Because the lapse of time denoted by *demain* 'tomorrow' is located after the moment of speech, the tense of the verb should be future as in (9a). In (9b) the verb *lisait* 'read' locates the event in the past, thus outside the temporal frame specified by *demain* 'tomorrow'. Accordingly, based on the definition given above, (9b) is a genuine tense violation. Anomalous verbs evoked a bi-phasic wave in the 450-550 ms interval following the onset of the critical word, with a negative maximum at posterior sites and a positive anterior peak. Given current knowledge of ERP components modulated by linguistic processes, it is hard to make sense of these effects, as they do not fit neatly any of the (E)LAN, N400 or P600 classes. Moreover, the Imparfait is typically used to describe a background eventuality (e.g. a state), against which a foreground event is described using the Passé Simple. The Imparfait in (9b) might introduce, on top of the tense violation itself, a background/foreground structure, with an expectation for a foreground event to be mentioned later in discourse. The Passé Simple *lut* 'read' would have been a more suitable candidate for realizing tense violations without such aspectual confounds.

Steinhauer and Ullman (2002) presented a group of English speakers with sentences like:

(10) a. Yesterday, I *sailed* Diane's boat to Boston.

b. *Yesterday, I *sail* Diane's boat to Boston.

In (10), *yesterday* specifies a past time frame for the occurrence of the main event. In (10b), *sail* locates the event in the present, thus outside the period denoted by *yesterday*. In the 400-900 ms time window following verb onset, tense violations elicited a consecutive LAN (400-500 ms) and P600 (600-900 ms), taken by the authors as signatures of morpho-syntactic processing. Regardless of whether *sail* was perceived as semantically or morphologically anomalous (recall the remarks above), the LAN observed by Steinhauer and Ullman (2002) indicates that semantic information, in the form of temporal constraints – or something functionally equivalent – set up by the adverb, is used by the system as early as 400 ms after verb onset.

The study reported below was aimed at obtaining a more complete picture of the time course of the neurophysiological effects of tense processing. We sought to refine the results of Steinhauer and Ullman (2002) using Dutch stimuli, in which the position of the verb in the sentence structure immediately follows that of the temporal adverb. Dutch syntax affords a semantically less spurious measurement of the effects of tense violations, as the arguments of the verb are yet to be processed when the tense violation occurs. This fact might also allow us to estimate more accurately the onset of the LAN and, as a consequence, to address issues such as to what extent temporal constraint satisfaction is implemented in the brain as a feed-forward process

(more on this later). Finally, we were interested in tracking more downstream ERP effects, such as those elicited by sentence-final words in narratives containing tense violations.

An ERP study of tense violations

Materials

The Dutch materials used in the experiment were 80 critical sentences, 40 correct sentences and 40 with tense violations, and 240 fillers. All sentences had the same structure: a past temporal adverb was followed by the main, simple past or present tensed verb, a subject NP and an object NP, in most cases modified by a PP. The following are examples of the stimuli used:

(S1) Afgelopen zondag *lakte* Vincent de kozijnen van zijn landhuis.

last Sunday painted Vincent the window-frames of his country-house.

‘Last Sunday Vincent painted the window frames of his country house.’

(S2) Afgelopen zondag *lakt* Vincent de kozijnen van zijn landhuis.*

last Sunday paints Vincent the window-frames of his country-house.

‘Last Sunday Vincent paints the window frames of his country house.’

The temporal adverbs were *vorige week* ‘last week’, *vorige maand* ‘last month’, *vorig jaar* ‘last year’, *vorige eeuw* ‘last century’ and *afgelopen N* ‘last N’ for each day of the week, each month and each season. Each adverb was used at most three times.

Eighty verbs were used, 40 regulars and 40 irregulars, 20 activities, 25 accomplishments and 35 achievements (Vendler, 1957). Mean length of the verbs (correct: $M=7.41$, $SD=2.56$; violation: $M=6.79$, $SD=2.25$) and raw frequency (correct: $M=6898.34$, $SD=34220.5$; violation: $M=7174.91$, $SD=44939.35$) were normed using the CELEX corpus (Baayen, Piepenbrock, & Gulikers, 1996), ensuring that there were no differences between conditions (T -tests, $P>0.9$ in all comparisons). Non-critical lexical items, including sentence-final words, and sentence length were identical across conditions. Fillers were 160 grammatically well-formed sentences containing the prepositions *after* and *before* and 80 verb-adverb sentences (40 correct, 40 tense violations) similar to critical items but constructed using different lexical material. Two test versions were constructed, consisting of pseudo-randomized lists of critical and filler items.

Participants

Twenty five students participated in the experiment. Of these, one was left out of the final analysis due to a high number ($>20\%$) of trials contaminated by artifacts. The remaining 24 participants (mean age 24.6, 14 female) had no history of neurological, psychiatric or cognitive disorders. Participants were selected from the database of the F.C. Donders Centre for Cognitive Neuroimaging at the Radboud University Nijmegen. They received €6 per hour or course credits for participating in the experiment.

Procedure

After applying the electrodes, participants were conducted into the experimental room and were asked to sit in front of a video monitor. The stimuli were presented on the screen word-by-word (300 ms word duration, 300 ms blank between words, white on black background), followed by a fixation cross lasting for 1500 ms. Participants were instructed to read each sentence carefully and to blink or move only when the fixation cross was shown. The experiment took about 50 minutes to be completed and was divided into 4 blocks of 80 trials each.

Recording

The EEG/EOG was recorded from 32 sintered Ag/AgCl electrodes. Two electrodes were placed at the outer canthi of the left and right eyes. One electrode below the left eye monitored vertical eye movements (e.g. blinks). The remaining 29 electrodes were arranged according to American Electrophysiological Society conventions at the following locations: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, TP10, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2. The left mastoid electrode served as the true reference. All electrodes were re-referenced off-line to a linked mastoid. Electrode impedance was kept below 5 k Ω throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with the

following settings: 500 Hz sampling rate, a low pass filter at 70 Hz and a 10 s time constant.

Data Analysis

Data were analysed using FieldTrip, a MATLAB package for processing EEG/MEG data. Several transforms were applied to each participant's data-set. Segments corresponding to the verb and the sentence-final word were extracted from the EEG with an interval of 200 ms before and 1000 ms after stimulus onset. Baseline correction used the 200 ms pre-stimulus interval. Two FieldTrip procedures were applied for artifact rejection. The first rejected all trials containing activity exceeding a threshold of $\pm 100\mu\text{V}$. The second algorithm discarded trials contaminated with eye movements or blinks based on thresholding the z-transformed value of the raw data in the EOG channels, preprocessed using a band-pass filter of 1-15 Hz. A 30 Hz low-pass filter was applied to the segmented, artifact-free data. ERPs were obtained for each participant by averaging over trials in each experimental condition. A 5 Hz low-pass digital filter was used to produce the waveforms shown in Figures 1-2.

Topographical plots and statistical analyses used the 30 Hz low-pass filtered data. Statistical analyses were based on a non-parametric randomization procedure (Maris, 2004; Maris & Oostenveld, 2007) which took as input mean amplitude values in each condition in time bins of 100 ms, starting from the onset of the relevant word and ending 1000 ms after, and produced as output a cluster of (1 to 28) electrodes in

which the difference between the conditions was significant in each time bin, the sum of T -statistics in that cluster and Monte Carlo estimates of P -values.

Results

INSERT FIGURES 1 AND 2 APPROXIMATELY HERE

INSERT TABLE 1 APPROXIMATELY HERE

A visual inspection of ERP waveforms elicited by the verb (Figure 1b) reveals a negative deflection peaking around 100 ms after verb onset, followed by a positive shift with a trough around 200 ms. The amplitude of these two components, often referred to as N1 (or N100) and P2 (or P200) respectively, appears not to be affected by the experimental manipulation: there are no electrode clusters between 0 and 200 ms after verb onset at which the mean amplitude difference between tense violations and correct sentences is significant (Table 1). The P2 is followed by a negative shift, larger for tense violations over left-anterior scalp sites, starting around 200 ms from verb onset and lasting for approximately 200 ms (Figure 1b). There are clusters of electrodes in which the negativity is significant between 200 and 400 ms (Table 1, Figure 1a), and marginally significant between 400 and 500 ms. This effect is an instance of left anterior negativity (LAN). Around 600 ms following verb onset, the waveforms are characterized by a bi-phasic (positive-negative) response (Figure 1b), which, however, is not different between conditions in either the 500-600 ms or the 600-700 ms time bin (Table 1). A positive shift, larger for tense violations over right-

posterior electrodes sets on at about 700 ms and lasts for the entire epoch (Figure 1): only marginally significant positive clusters are found between 700 and 1000 ms (Table 1). This effect can be taken as an instance of P600 (Osterhout & Holcomb, 1992; Hagoort et al., 1993).

The waveforms evoked by sentence-final words show similar N1-P2 complexes (Figure 2b). The amplitude of these early and largely endogenous components is not affected by verb tense (Table 1). The P2 is followed by a negative shift, peaking at about 400 ms following the onset of the sentence-final word (Figure 2b). There are no significant clusters between 200 and 400 ms, but the negativity becomes significantly larger for tense violations between 400 and 700 ms (Figure 2, Table 1). The effect bears some superficial resemblance with the N400, in particular as far as its polarity and distribution are concerned (Kutas & Hillyard, 1980). However, it has a later maximum and is more sustained compared to the N400. In what follows, I will refer to this effect as sentence-final negativity (SFN).

Tense at the Syntax-Semantics Interface

The link between linguistic theory and processing data is notoriously problematic (Poeppel & Embick, 2005). In particular, the notion that behavioral and brain data can constrain the form of linguistic theories – in somewhat more traditional terms: that the theory of ‘competence’ can be shaped by ‘performance’ data – is far from accepted

(Jackendoff, 2002). Following Jackendoff (2007), I assume that a linguistic theory can be preferred over its competitors also based on its capacity to account for data on language processing and acquisition, at a fairly abstract level of analysis (Marr, 1982; Baggio & van Lambalgen, 2007). Below I will sketch a theory of tense at the syntax-semantics interface that meets, at least to some extent, this requirement. In particular, the model is consistent with the observation that the constraints set up by a temporal adverb are used to process a main verb as early as 200-300 ms after the onset of the latter.

Tense, parsing and comprehension

An important missing ingredient from the semantic analysis of temporal expressions presented above is syntax. Here I will adopt a lexicalist parser developed by Vosse & Kempen (2000) and applied by Hagoort (2003) to account for syntax-related ERP effects. The model assumes that syntactic information is stored in the mental lexicon as three-tiered unordered trees, or *syntactic frames*, which specify the admissible syntactic environment of each word. The syntactic frames associated with *last*, *Sunday*, *Vincent* and *paint* are shown in (11), where *on* is assumed to be silent in examples (1a-b). The top tier of each frame is constituted by a single root node which contains phrasal information (NP, S etc.) and which dominates a set of *functional nodes* in the second tier. For instance, an S node dominates four functional nodes: subject, head, direct object and modifier; an NP node will have at most three

functional nodes: determiner (if it is not associated with a proper name, at least in some languages), head and modifier; and so forth. The third tier again contains phrasal nodes, called *foot nodes*, to which the root node of another syntactic frame can be attached.

(11) [figure provided separately as “(11).pdf”; my preference would be to typeset the contents of the figure (syntactic trees) using the style of the many text, rather than pasting it as a figure.]

Parsing consists in (i) checking agreement features of words such as number, gender, tense etc., (ii) checking word order, and (iii), only if these checks are passed, unifying frames with identical root and foot nodes. Subscripts in the labels of root and foot nodes are used to indicate unification links in (11). Unifying root and foot nodes with the same subscripts yields the syntactic structure of (1) up to and including the verb *painted/paints*. Vosse and Kempen (2000) assume that functional nodes carry features upon which the agreement checking (i) is based. For example, the subject node of *paints* carries the features *Case=nominative*, *Person=third* and *Number=singular*. These are compatible with the features carried by the root node of the frame associated with *Vincent*. Therefore, the check yields a positive result and the two frames are unified. Tense could be treated analogously using an additional level of features in the frames of verbs and temporal prepositions. However, the resulting feature structure would then have to be interfaced with the semantics of verbs and

temporal adverbs to ensure that the value of the *Tense* feature is consistent with that of the temporal variables in the semantic structures (3)-(5). This implies that at least three (sub)levels of representation have to be coordinated: the syntactic frame, the feature structure, and the semantic structure.

A more parsimonious option is to handle the check for temporal agreement entirely within the semantics. Consider (1a). Suppose that the parser has constructed a syntactic representation for *last Sunday* by unifying the NP foot and root nodes of *on* (which is silent) and *Sunday*, and the AP foot and root nodes of *Sunday* and *last*. Suppose, moreover, that the processor has constructed a semantic representation for the adverb *last Sunday* by forcing the query $HoldsAt(fCPSu, t)$ to succeed. This leads to a model in which the antecedent of (5) holds; this model satisfies the constraint $s < now$. Now, when *painted* is encountered, the syntactic frame of *paint* (which is tenseless) is retrieved from memory. Now the PP root and foot nodes of *last Sunday* and *paint* are unified as soon as the agreement check is passed. This is carried out entirely within the semantics by simply updating the initial discourse model according to the integrity constraints (3). The unification $s = t$ can be made to succeed, so the temporal agreement check is passed, and the two frames are unified. Consider now (1b). In this case, when *paints* is encountered, the initial model is updated according to (4). The adverb satisfies the constraint $s < now$, but the verb satisfies $t = now$, which entails that s and t cannot be unified. The agreement check (i) is not passed, which results in a failure to unify the syntactic frames of *last Sunday* and *paints*.

Toward a functional account of the LAN

Negative shifts in the ERPs different from the N400 have been reported in several studies. Some of these effects are usually referred to as (early) left anterior negativities or (E)LANs. The onset of (E)LANs ranges between 150 and 300 ms as a function of stimulus characteristics (Friederici, 2002). ELANs (150-200 ms) have been observed in response to rapidly detectable violations of word-category. LANs (300-500 ms) are elicited by morphological anomalies such as case, number and gender mismatches (Friederici, 2002; Hagoort, 2003). As the data show, tense violations result in LANs too, setting in between 200 and 300 ms from verb onset.

The theory of tense at the syntax-semantics interface presented above seems to be capable of accounting for the early response to tense violations. One key observation in this regard is that an effect which sets in between 200 and 300 ms after the onset of the stimulus arguably does not reflect feed-back or recurrent computations. Rather, it is more likely related to the disruption of a largely automatic, feed-forward spread of activation from sensory areas (in this case, visual cortex), through brain regions subserving semantic memory (i.e. left temporal cortex), towards anterior language areas (i.e. left frontal cortex) (Hagoort, 2005). Hagoort (2003) suggested that (E)LANs might reflect the failure to bind two syntactic frames as a result of a negative outcome of the agreement check or of a failure to find two matching root and foot nodes. The LAN observed in the ERP experiment reported here can be taken as

reflecting the failure in the tense agreement check – more precisely, a failure to simultaneously solve the temporal constraints set up by the adverb and the verb. This account appears to be consistent with the time course of the LAN, for it only requires that the constraint $t=now$, associated with the tense suffix of the verb *paints* (or *lakt* in the Dutch case), is active in semantic memory. Given the constraints set up by the adverb in the preceding processing stages, it follows that $s=t$ (a condition of temporal coherence within the sentence) cannot be satisfied. Note that this account does not require (i) that the full meaning (assuming there is anything like that) of the verb *paints* is reconstructed from semantic memory, for only the semantic contribution of the tense suffix ‘-s’ is necessary, or (ii) that any form of syntactic structure assembly takes place, for the constraint satisfaction stage actually precedes the unification of syntactic frames.

Although in-depth experimentation and modeling are still lacking, the temporal profile of the LAN can be taken to suggest that checking the satisfiability of a set of temporal constraints – or perhaps, if one can see a difference there, just detecting an inconsistency in such a set – might have a feed-forward neural implementation. This can be contrasted with processes such as the (re-)computation of a discourse model, which require a recurrent network architecture (Hitzler, Hölldobler, & Seda, 2004; Stenning & van Lambalgen, 2005; Baggio & van Lambalgen, 2007) and gives rise to qualitatively different neurophysiological effects (Baggio, van Lambalgen, &

Hagoort, 2008). This observation may constitute a starting point for an explanation of the SFN.

Towards a functional account of the SFN

A number of studies have reported negative shifts in the ERPs in response to sentencefinal words in syntactically ill-formed sentences, even when the anomalous word does not occupy the sentence-final position (Osterhout & Holcomb, 1992; Hagoort et al., 1993; Osterhout & Holcomb, 1993; Osterhout & Mobley, 1995; McKinnon & Osterhout, 1996; Osterhout, 1997; Osterhout, Bersick, & McLaughlin, 1997). Osterhout and Nicol (1999) reported negativities in response to sentence-final words preceded by semantic as well as syntactic violations. Moreover, at some electrode sites, the SFN elicited by doubly (syntactically and semantically) anomalous sentences was approximately an additive function of the SFNs evoked by syntactic and semantic violations.

Osterhout and Holcomb (1992) suggested that the SFN may be an electrophysiological marker of either (i) the perceived ungrammaticality of the sentence, or (ii) the system's effort to find an acceptable syntactic structure for the sentence, or (iii) the 'message-level' consequences of the sentence-internal violation. As for the SFN elicited by tense violations, (i) seems to imply either that (a) the incorrect use of tense was not perceived when the verb was encountered, but only at the sentence-final word, or that (b) the ill-formedness of the *sentence* (as opposed to

that of the VP) was perceived at the sentence-final word, as only then it can be concluded that the entire sentence, which consists of a single VP, is ungrammatical. Now, both these implications seem untenable. For one, the LAN elicited by the verb indicates that the violation *was* in some sense perceived, which rules out (a). Second, any sentence containing an ungrammatical phrase is itself ungrammatical, which makes (b) untestable. Hypothesis (ii) implies that, at the sentence-final word, the system computes a plausible syntactic representation for the anomalous sentence (1b), eventually unifying the PP root and foot nodes of ‘last Sunday’ and ‘paint’. However, computing a new syntactic representation does not solve the sentence-internal problem, which is semantic in nature. Either syntactic unification does away with the agreement check at the sentence-final stage, in which case the unification of frames does not guarantee that tense and other agreement features are consistent across the representation, or the agreement check is still required, in which case it is the semantic constraints that must be readjusted.

This brings us to (iii), according to which the SFN would reflect either (a) the disruption of a process involved in computing a model of the sentence or (b) the attempt made by the system to compute a model in which all constraints are satisfied. One key feature of the Event Calculus combined with Constraint Logic Programming is that satisfying a set of constraints actually *produces* a discourse model verifying the linguistic material given as input (van Lambalgen & Hamm, 2004). In this framework, the LAN can be taken as a physiological index of the failure to unify *s*

and t (that is, to add the constraint $s=t$ to the unification space), and thereby of the disruption of the processes which would otherwise have led to a discourse model. This rules out (a). According to (b), the SFN would reflect the readjustment of constraints set up by the verb required to make the set of constraints satisfiable as a whole. In semantic terms, this process can be characterized as, for instance, forcing the integrity constraints (3) in the present tense case (1b), yielding a ‘narrative present’ reading of the sentence. In terms of neural implementation, reworking a set of constraints to produce the desired output requires a recurrent architecture. Feedback processes seem consistent with the temporal profile of the SFN (400-700 ms), and complete the picture sketched above where the LAN (200-400 ms) reflects a feed-forward computation instead.

Summary and Conclusions

The aim of the present paper was to take a few initial steps towards an integrated account of temporal reference in terms of linguistic structure and neural processes. I have tried to suggest that the bulk of linguistic computation in the brain, as far as temporal reference is concerned, is semantic. I argued that verb tense can be said to be used correctly or incorrectly only based on a semantic criterion, which can be captured in terms of temporal constraints using logical systems like the Event Calculus. Further on, I outlined an integrated model of parsing and comprehension in

which the computation of temporal reference is entirely consigned to the semantic processor and acts as a preliminary stage of syntactic structure building. Under this view, the ERP data reported here, and the LAN in particular, suggest that the brain accesses and evaluates semantic information as early as 200-300 ms following word onset.

Author Note

I wish to thank Oliver Bott, Emmanuel Chemla, Bart Geurts, Peter Hagoort, Fritz Hamm, Wolfgang Klein, Peter Indefrey, Michiel van Lambalgen and Leah Roberts for many valuable comments on the issues discussed in this paper. I am grateful to the Netherlands Organization for Scientific Research (NWO) for support under grant 051.04.040-NWOCOG/04-19.

References

- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes, 18*, 405-430.
- Baayen, R., Piepenbrock, R., & Gulikers, L. (1996). *CELEX2*. Linguistic Data Consortium, Philadelphia.
- Baggio, G., & van Lambalgen, M. (2007). The processing consequences of the imperfective paradox. *Journal of Semantics, 24*, 307-330.
- Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language, 59*, 36-53.
- Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.
- Fonteneau, E., Frauenfelder, U., & Rizzi, L. (1998). On the contribution of ERPs to the study of language comprehension. *Bulletin suisse de linguistique appliquée, 68*, 111-124.
- Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78-84.
- Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
- Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage, 20*, S18-S29.

- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9, 416-423.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.
- Hamm, F., Kamp, H., & van Lambalgen, M. (2006). There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 22, 1-40.
- Hitzler, P., Hölldobler, S., & Seda, A. (2004). Logic programs and connectionist networks. *Journal of Applied Logic*, 3, 245-272.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, 1146, 2-22.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, 11, 539-550.
- Larson, R., & Segal, G. (1995). *Knowledge of meaning*. Cambridge, MA: MIT Press.
- Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41, 142-151.

- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177-190.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *11*, 495-523.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494-522.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, *25*, 273-285.
- Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.
- Osterhout, L., & Holcomb, P. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413-437.
- Osterhout, L., & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739-773.

- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 283-317.
- Parsons, T. (1990). *Events in the semantics of English*. Cambridge, MA: MIT Press.
- Partee, B. (1973). Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy, 70*, 601-609.
- Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103-118). Lawrence Erlbaum Associates.
- Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language, 83*, 62-65.
- Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science, 29*, 919-960.
- van Lambalgen, M., & Hamm, F. (2004). *The proper treatment of events*. Oxford: Blackwell.
- Vendler, Z. (1957). Verbs and times. *The Philosophical Review, 66*, 143-160.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75*, 105-143.

[Table 1 provided as separate file “Table1.pdf”]

TABLE 1. Summary of cluster-based *T*-statistics for the ERP data. Tense violations and correct sentences are compared at the verb and at the sentence-final word in time bins of 100 ms starting from word onset. The first significant effects occurred at 200-300 ms. Empty cells denote the absence of (marginally) significant clusters.

[Figure 1 provided as separate file “Figure1.pdf”]

FIGURE 1. (a) Grand-average ($N=24$) topographies displaying the mean amplitude difference between the ERPs evoked by the verb in tense violations and in correct sentences. (b) Grand-average ($N=24$) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in tense violations and correct sentences. Negative values are plotted upwards.

[Figure 2 provided as separate file “Figure2.pdf”]

FIGURE 2. (a) Grand-average ($N=24$) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final word in tense violations and correct sentences.

(b) Grand-average ($N=24$) ERP waveforms from frontal, central and parietal electrode

sites time locked to the onset (0 ms) of the sentence-final word in tense violations and correct sentences. Negative values are plotted upwards.

THE PROCESSING CONSEQUENCES OF THE IMPERFECTIVE PARADOX

GIOSUÈ BAGGIO AND MICHIEL VAN LAMBALGEN

ABSTRACT. In this paper we present a semantic analysis of the imperfective paradox based on the Event Calculus, a planning formalism characterizing a class of models which can be computed by connectionist networks. Further, we report the results of a questionnaire which support our semantic theory by showing that different aspectual classes of progressive VPs give rise to different entailment patterns. A processing model is then outlined, combining the semantic analysis with the principle of immediacy as established in psycholinguistics. Finally, we derive predictions concerning the ERP correlates of the processes described by the theory.

1. INTRODUCTION

Recently an increasing number of studies has brought experimental data to bear on specific semantic hypotheses.¹ However, theoretical semantics and psycholinguistics have reached their most important results independently, even when the issues at stake could have been addressed more completely by joining the efforts. One example in this regard is the study of discourse-based inferences, where logic and formal semantics, which have a lot to say about entailment and related notions, were often neglected in psycholinguistics.²

The assumption behind this paper is that there exists a relatively unexplored territory in which the two disciplines can interact productively. Here we consider a small portion of such territory: the imperfective paradox and its processing consequences. First we provide a minimal methodological background against which semantic theory can be combined with models of language processing and we introduce the semantic paradox which we shall work with. In section 2 we present a semantic analysis of the paradox based on the Event Calculus [van Lambalgen and Hamm, 2004], a planning formalism characterizing a class of models which can be computed by connectionist networks, as explained in section 4.3. In section 3 we report the results of a questionnaire showing that different aspectual classes of progressive VPs give rise to different entailment patterns. In section 4 a processing model is outlined, combining the semantic analysis with the principle of immediacy as established in psycholinguistics. The resulting ‘recomputation hypothesis’ is used to derive predictions concerning the ERP correlates of the processes described by the theory. We conclude with some remarks on extending the proposed approach to other linguistic phenomena.

1.1. Methodological background. From a semantic analysis of a given linguistic structure, for instance using possible worlds semantics or Discourse Representation Theory (DRT), it is usually not possible to derive predictions concerning the complexity and the time-course of the relevant computations. Although there may be cases in which processing hypotheses can be formulated on the basis of the semantic theory alone [Geurts and van der Slik, 2005, McMillan et al., 2005], this method might not work in general. The reason is that processing largely depends upon the particular algorithms and mechanisms that, in a physical system such as the human brain, compute the linguistic structures posited by the theory – or some functionally equivalent structure. Because semantics typically does not describe algorithms and neural mechanisms (nor perhaps it should), there is no direct way to relate semantic theory to what is observed in actual language processing experiments. How can we make semantics part of a processing theory?

¹See for instance the work on quantifiers by [Geurts and van der Slik, 2005] and [McMillan et al., 2005].

²See [Cook et al., 2001] and [Frank et al., 2003] among others.

The solution is to develop a theoretical framework which allows semantic structures to be explicitly related to processing algorithms inspired by the available psycholinguistic evidence and ultimately to mechanistic models of the kind investigated in neuroscience. Marr's [Marr and Poggio, 1976, Marr, 1982] tripartite scheme for the analysis of cognitive systems seems a reasonable choice in this regard [Baggio et al., 2007].

[Marr, 1982] argued that information processing systems should be understood at three, nearly independent levels of analysis. The first contains a description of the computations to be performed by the system and, more precisely, a characterization of the *goals* that have to be attained in order to solve the information processing problem. For instance, a sentence *S* can be seen as introducing a specific goal, namely the construction of a model making *S* true. Accordingly, in the semantic analysis presented below, the tense and aspect of VPs like the progressive are treated as instructions to update the current discourse model so as to achieve that goal. *We regard it as the task of semantics to describe information processing goals and update instructions.* The actual processing steps are described at the intermediate level, where constraint satisfaction algorithms and processing principles such as immediacy and incrementality are combined. Marr would complete the picture with a third level, at which the neural mechanisms subserving the computations are described. We will discuss issues relevant for the computational level in sections 2 and 3, for the algorithmic level in sections 4.1 and 4.2, and for the level of the neural implementation in section 4.3.

1.2. The imperfective paradox. [Vendler, 1957] famously classified VPs as states ('know', 'love', etc.), activities ('run', 'write', etc.), accomplishments ('write a letter', 'bake a cake', etc.), achievements ('finish', 'reach', etc.) and points ('flash', 'hop', etc.). We are interested in the inferences licensed by sentences containing activities and accomplishments in the past progressive. The following example involves the accomplishment 'write a letter':

- (1) The girl was writing a letter when her friend spilled coffee on the tablecloth.

From (1) the reader would typically conclude that, barring unforeseen circumstances, the girl attained the desired goal and would thus assent to the statement 'The girl has written a letter'. Such inference is based on the assumption that spilling coffee on the tablecloth is usually *neutral* with respect to the writing activity. That is, it is not a typical immediate cause leading to the termination of the activity. One can easily imagine situations in which writing is temporarily interrupted or even terminated by the accident. However, as the data reported in section 3 will demonstrate, failing to explicitly mention a disabling condition in the discourse is sufficient to lead the reader to assume that there was no such obstacle to attaining the intended goal.

Goal state inferences are *non-monotonic*, that is they can be suppressed if the discourse describes an event which terminates the relevant activity:

- (2) The girl was writing a letter when her friend spilled coffee on the paper.

Assuming that writing was intended to occur on the same paper sheets on which coffee was spilled, the accident is sufficient to terminate the activity and it is therefore a *disabling* condition for obtaining a complete letter. Accordingly, on the basis of (2) the reader would assent to 'The girl has written no letter'.

One important observation is that suppression can obtain only with accomplishments, and not with activities. Since a sentence containing an activity in the past progressive, such as 'writing letters', does not involve a canonical goal, it is interpreted as entailing the truth of 'The girl has written one or more letters' regardless of the consequences of the second event on the writing activity:

- (3) The girl was writing letters when her friend spilled coffee on the tablecloth.
 (4) The girl was writing letters when her friend spilled coffee on the paper.

Upon reflection, there is something paradoxical about examples (1) and (2). Whereas it belongs to the meaning of the accomplishment 'writing a letter' that the writing activity is directed toward the consequent state of a complete letter, the actual occurrence of that

consequent state can be denied without contradiction. But how can a seemingly essential component of the meaning be denied without affecting the meaning itself? This is known as the ‘imperfective paradox’. The semantic literature is replete with attempted resolutions of the paradox, ranging from explaining the problem away [Michaelis, 2001] to invocations of possible worlds [Dowty, 1979, Landman, 1992, de Swart, 1998]. It is impossible to review all proposed solutions here; we will focus on some representative cases. Possible worlds solutions are based on the idea that

the progressive picks out a stage of a process/event which, if it does not continue in the real world, has a reasonable chance of continuing in some other possible world [de Swart, 1998, p. 355].

They differ however in the (largely informal) descriptions of the possible worlds used. For example, [Dowty, 1979] would claim that the following are equivalent:

- a. ‘The girl is writing a letter’ is true in the actual world;
- b. ‘The girl will have written a letter’ is true in all so-called ‘inertia worlds’, worlds which are identical with the present world until ‘now’, but then continue in a way most compatible with the history of the world until ‘now’.

These approaches are intensional in the formal sense of using possible worlds. In fact, most authors (but not all) would agree that the progressive creates an intensional context: even though the accident in (2) may have terminated the writing activity at a stage in which it was unclear whether the girl was writing a letter or, say, a diary entry, still only one of

- (5) The girl was writing a letter.
- (6) The girl was writing a diary entry.

is true of the situation described by (2). Explicitly denying that the progressive creates an intensional context, Michaelis argues that³

the Progressive sentence ‘She is drawing a circle’ denotes a state which is a subpart not of the accomplishment type *She-draw a circle* but of the activity type which is entailed by the causal representation of the accomplishment type. Since this activity can be identified with the preparatory activity that circle drawing entails, circle drawing can in principle be distinguished from square drawing etc. within the narrow window afforded by the Progressive construal [and] does not require access to culmination points either in this world or a possible world. [Michaelis, 2001, p. 38]

We find this questionable, for without access to a person’s intention it may be very hard to tell initially whether she is drawing a circle or a square, writing a letter or a diary entry. But that person’s intention in performing an activity is characterized precisely by the associated consequent state, even though the latter cannot yet be inferred from the available data.

Here the Event Calculus will come to our rescue, because the notion of goal or intention is built into the semantics from the start. In particular, the meaning of a VP is represented by a *scenario* which describes a plan for reaching the goal. However, unlike approaches such as [Parsons, 1990], where one quantifies existentially over events, the scenario is a universal theory and does not posit the occurrence of the intended consequences. Although the plan is appropriate for that purpose, attaining the goal is guaranteed only in a *minimal model* (in which no unforeseen obstacles occur) of the scenario plus the axioms of the Event Calculus. The meaning of an accomplishment (as embodied in the scenario) involves a culminating event *type*, which therefore must exist; but no existential claims are made concerning the corresponding event *token*, which, as in example (2), may also fail to occur. Type and token are handled by different mechanisms. These notions form the basis of our analysis of the imperfective paradox in the Event Calculus, to which we now turn.

³Replace ‘drawing a circle’ with ‘writing a letter’ and ‘drawing a square’ with ‘writing a diary entry’ in Michaelis’ examples to see the connection with our examples.

2. SEMANTIC ANALYSIS

The reference to goal states in the preceding section suggests that a semantic analysis of the progressive can be based on a planning formalism which is able to talk about goals and actions, and which includes a theory of causality together with a principle of inertia. Such formalism is presented in [van Lambalgen and Hamm, 2004]. It consists of an Event Calculus which has found applications in robotics, reformulated using the computational machinery of Constraint Logic Programming.

The reader may wonder why planning can provide a source of inspiration to linguistics. The reason can be found in the nature of planning computations, which typically proceed as follows. First a goal is specified, which may be an action to be performed at a particular location (for example, pick up outgoing mail in an office building). Next a plan is computed, that is a sequence of actions to get to the required location, derived by backward chaining from the goal to obtain a sequence of subgoals, the last one of which can be executed in the agent's initial position and state. Planning requires a situation model (including a map of the building, a causal theory of the agent's actions, a specification of values of variables such as 'door open/closed', the agent's initial position and state, a record of its past and current actions), a repertoire of possible actions ('follow wall', 'go through door') and observations ('door open/closed'). While the agent executes the plan, it also registers its observations and actions in the situation model; knowledge of its actions may be important for the agent to estimate its current position. Particularly relevant for our purposes is that a plan might have to be recomputed in mid-course when the situation model is updated due to new observations (for instance, a closed door which was expected to be open on the basis of the initial model, or a wrong estimate of the agent's current position). Later on, in section 4.2, we shall see how recomputation relates to the imperfective paradox.

This short description should be sufficient to enable the reader to see the connection with language processing. The comprehender starts with an initial discourse model, in which an incoming sentence or clause must be integrated. Suppose the main verb of the sentence is non-stative, for instance an activity.⁴ If the sentence is in one of the simple tenses, it is unpacked in the relevant action and its participants, and the discourse model is updated accordingly. This is the analogue of updating the situation model with representations of individuals and actions in planning. In more complex cases, such as sentences involving accomplishments like (1) and (2) above, the VP will be taken to express the existence of a goal-directed plan. If, on the contrary, the main verb of the sentence is stative, the sentence can be viewed as analogous to an observation report, and the discourse model is updated with a property.

2.1. The Event Calculus. The Event Calculus is a planning formalism which allows one to talk about actions, goals and causal relations in the world. Its main function is to return a plan given a goal, the initial state and causal relations. Formally, the Event Calculus is a many-sorted predicate logic. It has two different sorts for events, viewed either perfectly or imperfectly. The former are called *event types* and are symbolized by e, e', \dots, e_0, \dots . The latter are called *fluents*⁵ and are symbolized by f, f', \dots, f_0, \dots . One may think of event types as action types, such as for example 'break' or 'ignite'; fluents can be thought of as time-dependent properties, for example 'being broken' or 'writing'; the time parameter in fluents is implicit, but they can have further parameters (for instance, for the subject of 'writing'). The real distinction between event types and fluents comes from the different roles they play in the axioms of the Event Calculus.⁶ The universe must also contain sorts for individuals ('the girl'), real numbers interpreted as instants of time, and various other real quantities (e.g. position, velocity, degree of some quality).

⁴For simplicity the reader may think of the English case.

⁵Newton's name for time-dependent variables.

⁶For which see [van Lambalgen and Hamm, 2004, pp. 39-41].

The primitive predicates comprise the minimum necessary to talk about two forms of causality, instantaneous (as in a collision) and continuous (as when a force is acting):

- (7) a. *Initially*(f): ‘fluent f holds at the beginning of the discourse’
 b. *Happens*(e, t): ‘event type e has a token at time t ’
 c. *Initiates*(e, f, t): ‘the effect of event type e at time t is the initiation of f ’
 d. *Terminates*(e, f, t): ‘the effect of event type e at time t is the termination of f ’
 e. *Trajectory*(f_1, t, f_2, d): ‘if f_1 holds from t until $t + d$, then at $t + d$, f_2 holds’
 f. *Releases*(e, f, t): ‘event type e releases the continuously-varying fluent f ’⁷
 g. *HoldsAt*(f, t): ‘fluent f holds at at time t ’

While (7a-d) are the predicates required by instantaneous change, (7e-f) are used to model continuous change. The meaning of these primitive predicates is determined by the axioms of the Event Calculus.

There are three key notions which will be used in the semantic analysis. The first is that of *scenario*.⁸ Whereas the axiom system EC provides a macro-theory of causality, scenarios provide micro-theories stating the specific causal relations obtaining in a given situation. Scenarios can be used to describe the causal environment of actions and events at the level of granularity typically required by natural language, such as for instance writing a letter or drawing a circle. Scenarios can be taken to represent, in a tenseless fashion, the meaning of VPs. For instance, the scenario for the accomplishment ‘writing a letter’ specifies that the writing activity is causally related to the amount of letter produced, that the termination of the activity is temporally contiguous to the completion of the letter, and so on.

The second notion is that of *integrity constraint*. As we said above, we regard a sentence as introducing an information processing goal (‘Make S true’) to be achieved by updating the current discourse model or constructing a new one. Integrity constraints regiment such updates. They can take the form either of obligations ‘ $?\varphi$ succeeds’, requesting an update of the discourse model satisfying φ , or of prohibitions ‘ $?\varphi$ fails’, blocking an update of the discourse model satisfying φ . While scenarios describe the meaning of VPs in a tenseless fashion, integrity constraints specify the contribution of tense, aspect and temporal adjuncts to the semantics of VPs.

The third notion is that of *minimal model*. The axioms of EC constitute a correct theory of causality if and only if the following two requirements are satisfied:

1. The discourse model contains only those occurrences of events forced to be there by the discourse and the axioms;
2. The interpretation of the primitive predicates is as small as is consistent with the discourse and the axioms.

These two requirements define minimal models. They imply that no unforeseen events are allowed to happen and that all causal influences are as expected. The choice to work with minimal models instead of all models leads to *non-monotonicity* in discourse interpretation: adding a new sentence or clause to the discourse may invalidate a conclusion derived from the initial model. In [van Lambalgen and Hamm, 2004] it is argued that it is precisely the possibility to retract previously inferred conclusions which allows a rigorous treatment of the imperfective paradox. The most important meta-theorem about the EC formalism is that minimal models exist and can be computed efficiently [van Lambalgen and Hamm, 2004, Stenning and van Lambalgen, 2005, Stenning and van Lambalgen, 2007]. Further, minimal models can be computed (or approximated, depending on the expressiveness of the logical language) by connectionist neural networks – a topic to which we return in section 4.3.

⁷The predicate *Releases* is used to reconcile the two notions of causality: while instantaneous change leads to one form of inertia, where properties do not change their value between the occurrences of two events, continuous change requires that variable quantities may change their value without concomitant occurrences of events. The solution of this conflict is to exempt, by means of the predicate *Releases*, those properties which we want to vary continuously from the inertia of the first form of causation.

⁸See the related notions of ‘frame’ and ‘script’ in the classical papers [Minsky, 1977] and [van Dijk, 1980]. See also [Schank and Abelson, 1977] for a more explicit connection with planning.

2.2. The semantics of the progressive. The semantics of the VP ‘was writing a letter’ can be decomposed into the tenseless lexical meaning of ‘write a letter’ and the temporal and aspectual contribution of the past progressive. The lexical meaning is represented by the following scenario:⁹

- (8) a. *Initially*(*letter*(*a*))
- b. *Initiates*(*start*, *write*, *t*)
- c. *Initiates*(*finish*, *letter*(*c*), *t*)
- d. *Terminates*(*finish*, *write*, *t*)
- e. *HoldsAt*(*write*, *t*) \wedge *HoldsAt*(*letter*(*c*), *t*) \rightarrow *Happens*(*finish*, *t*)
- f. *Releases*(*start*, *letter*(*x*), *t*)
- g. *HoldsAt*(*letter*(*x*), *t*) \rightarrow *Trajectory*(*write*, *t*, *letter*(*x* + *g*(*d*)), *d*)

Here *write* is the activity fluent, *letter*(*x*) is a parametrized fluent for the completion stage *x* of the letter, *a* is a constant for the stage at which writing is initiated, *c* is a constant for the stage at which the letter is considered finished, and *letter*(*c*) is the fluent for the goal state. The dynamics of the scenario is formalized by (8g), which says that, if the stage of completion of the letter at time *t* is *x*, then the writing activity, lasting from time *t* until *t* + *d*, will result in a letter whose stage of completion at time *t* + *d* is *x* + *g*(*d*), where *g* is a monotone increasing time-dependent real-valued function relating the activity to the completion stage.¹⁰ Accomplishments are distinguished from activities by the statements regimenting the behavior of the goal state, here (8c) and (8e): because no canonical goal is involved in activities like ‘write letters’, their scenario will not contain these clauses.

The contribution of the past progressive is represented by the integrity constraint

- (9) *?HoldsAt*(*write*, *R*) \wedge *R* < *S* *succeeds*

which forces an update of the discourse model such that the activity fluent *write* holds at the reference time *R*, located in the past of the moment of speech *S* [Reichenbach, 1947]. As remarked above about examples (1)-(4), different inferences can be drawn from activities and accomplishments in the past progressive. This is accounted for in the Event Calculus by the fact that updating the initial model according to (9) leads to different minimal models depending on whether the scenario represents an activity or an accomplishment. If the VP is an accomplishment, Theorem 1 provides information on the non-monotonic (suppressible) character of the inferences licensed by accomplishments in the progressive. We mention this result to show that there is a formal underpinning of the theory, but we realize that for some readers these considerations will be too technical. An elucidation of the theorem will be provided below.

Theorem 1. *Let \mathcal{D} be a discourse consisting of scenario (8) given above. Suppose \mathcal{D} is extended to \mathcal{D}' so that the query $?HoldsAt(write, R) \wedge R < S$ succeeds in \mathcal{D}' . Suppose also $\lim_{t \rightarrow \infty} g(t) \geq c$. Then there is a unique minimal model of \mathcal{D}' , and in this model there is a time $t \geq R$ for which $HoldsAt(letter(c), t)$. By virtue of the stipulation $Letter(letter(c))$, there will be a letter at time t .*¹¹

If *write* holds at *R*, as required by (9), it must either hold initially or have been initiated. The latter requires an event *start* which initiated the writing activity. Since the starting event is not provided by discourse, we must assume that *write* holds initially. The clause (8g) states that the writing activity will increase the stage of completion of the letter. As time tends to infinity, the latter will be at least equal to *c* (the final completion stage). We have stipulated that a letter whose stage of completion is *c* is a finished letter. Hence, there must be a time at which the letter is considered finished. The writing activity will stop as soon as a complete letter is obtained, as is guaranteed by (8d-e). Notice that this holds for accomplishments, not for activities: if the VP is the activity ‘writing letters’, (8c) and (8e) are not part of the

⁹At present we have no algorithm to construct scenarios from more standard representations of lexical meaning.

¹⁰The way *g* is defined depends upon the particular VP (or, equivalently, the particular plan) at issue. For instance, ‘erasing the blackboard’ or ‘dismantling the engine’ would require a monotone *decreasing* function.

¹¹See [van Lambalgen and Hamm, 2004, p. 158-161] for a more rigorous statement and a computational proof.

scenario, and writing will continue also after one letter has been completed – in fact, it will continue indefinitely, if no terminating event is described.

We have just seen how the Event Calculus deals with the interpretation of VPs in the past progressive. We will now discuss the contribution of subordinate ‘when’ clauses to the meaning of (1)-(4). Because spilling coffee on the tablecloth usually does not terminate the writing activity, the dynamics of the scenario will lead to a complete letter in (1) and to an indefinite number of letters in (3). The argument mirrors the one given above as a gloss of Theorem 1. The situation is different for examples (2) and (4). Spilling coffee on the paper is typically sufficient to terminate the writing activity. This bit of world knowledge can be expressed by the following addition to scenario (8):

(10) *Terminates*(spill, write, t)

But what is more important here is the integrity constraint introduced by the ‘when’ clause

(11) *?Happens*(spill, R) \wedge R < S succeeds

which requires the accident to occur at R, while the writing activity was taking place. Since during the writing process there is no complete letter (yet), spilling coffee on the paper will terminate the activity *before* the letter is finished. Therefore, there will be no complete letter in the final discourse structure. As for (4), the theory is consistent with both the situation in which the writing activity was terminated before even a single letter had been completed and the case in which one or more letters were finished when the accident happened. Still, the proposed semantics does predict that readers would assent to ‘The girl has written a letter’ in (1), to ‘The girl has written no letter’ in (2) and to ‘The girl has written one or more letters’ in (3). This seems consistent with received semantic wisdom about entailments of activities and accomplishments in the progressive. Behavioral data supporting this claim are presented in the next section.

To summarize, the attainment of the goal state can be derived in a minimal model of a discourse containing an accomplishment in the progressive. However, the computation of discourse models is non-monotonic: when the initial minimal model is extended with a sentence or a clause describing an event which terminates the relevant activity (what we called a disabling condition), the derivation of the goal state is blocked. Non-monotonicity affords a neat solution of the imperfective paradox, as there is no contradiction in assuming that the goal state is a semantic component which is both essential to the meaning of the progressive VP (as an event type in the scenario) and suppressible on the basis of additional discourse information (as an event token in the minimal model).

3. BEHAVIORAL DATA

In the preceding sections we have argued that accomplishments and activities in the past progressive behave differently as regards entailment. An accomplishment such as ‘writing a letter’ implies that ‘The girl has written a letter’, provided that no obstacles are described in discourse. Failing to mention a disabling condition is sufficient to conclude that there is no such obstacle to attaining the goal. If however a disabling condition is introduced, then the accomplishment implies that ‘The girl has written no letter’. An activity like ‘writing letters’ implies that ‘The girl has written one or more letters’ regardless of further discourse information. Since disabling conditions affect the possibility of attaining a predefined goal, and such predefined goals are part of the meaning of accomplishments but not of activities, accomplishments will be sensitive to the presence of a disabling condition in the discourse context, whereas activities will not. The entailment questionnaire reported below aimed at testing this claim.

3.1. Subjects. Thirty two native speakers of Dutch (mean age 22.5, 27 female) completed the cloze probability test (see below) and thirty six (mean age 22.5, 28 female) the entailment questionnaire. Participants were selected from the database of the F.C. Donders Centre for Cognitive Neuroimaging in Nijmegen. The two sets of subjects were disjoint.

3.2. **Materials.** The set of Dutch materials used in the questionnaire included 160 test items. Each test item comprised two context sentences providing a neutral setting for the events described by critical sentences,¹² four critical sentences and two probe pairs:

Context sentences:

- (C) De deur van de woonkamer was gesloten. Binnen speelde de radio klassieke muziek.
The door of the living-room was_{PAST} closed_{PART}. Inside played_{PAST} the radio classical music.
(The door of the living room was closed_{PAST PART}. Inside the radio played_{PAST} classical music.)

Critical sentences:

- (S1) Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed morste.
The girl was_{PAST} letters on the to-write_{INF} when her friend coffee on the tablecloth spilled_{PAST}.
(The girl was writing_{PAST PROG} letters when her friend spilled_{PAST} coffee on the tablecloth.)
- (S2) Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier morste.
The girl was_{PAST} letters on the to-write_{INF} when her friend coffee on the paper spilled_{PAST}.
(The girl was writing_{PAST PROG} letters when her friend spilled_{PAST} coffee on the paper.)
- (S3) Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste.
The girl was_{PAST} a letter on the to-write_{INF} when her friend coffee on the tablecloth spilled_{PAST}.
(The girl was writing_{PAST PROG} a letter when her friend spilled_{PAST} coffee on the tablecloth.)
- (S4) Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste.
The girl was_{PAST} a letter on the to-write_{INF} when her friend coffee on the paper spilled_{PAST}.
(The girl was writing_{PAST PROG} a letter when her friend spilled_{PAST} coffee on the paper.)

Probe sentences:

- (P1) Het meisje heeft een of meer brieven geschreven.
The girl has_{PRES} one or more letters written_{PART}.
(The girl has written_{PRES PERF} one or more letters.)
 Het meisje heeft geen brief geschreven.
The girl has_{PRES} no letter written_{PART}.
(The girl has written_{PRES PERF} no letter.)
- (P2) Het meisje heeft een brief geschreven.
The girl has_{PRES} one letter written_{PART}.
(The girl has written_{PRES PERF} one letter.)
 Het meisje heeft geen brief geschreven.
The girl has_{PRES} no letter written_{PART}.
(The girl has written_{PRES PERF} no letter.)

Critical sentences were constructed manipulating the aspectual class of the VP in the past progressive (activity or accomplishment) and the causal type of the condition introduced by the subordinate 'when' clause (neutral or disabling with respect to the event described in the progressive clause). Activities and accomplishments differed only in the object NP: an indefinite ('een brief') was used for accomplishments and a bare plural ('brieven') for activities. Disabling and neutral conditions differed only in the prepositional or object NP, for instance 'papier' for the former and 'tafelkleed' for the latter. The distinction between neutral and disabling conditions was made according to the experimenters' judgment. The probes (P1) were used with activities, (P2) with accomplishments.

The following linguistic properties of critical sentences were normed. The average length and raw frequency of the differing words in the object NP of subordinate clauses were matched using the CELEX 2001 corpus.¹³ To determine the cloze probabilities of the verbs in the subordinate clauses, context items followed by one of the critical sentences with the final word blanked were presented to subjects. Participants were asked to fill in the blank space with the first word that came to their mind. Four versions (40 items per condition), randomized and balanced across conditions, were constructed. The results show that mean cloze probabilities are not different across conditions. This was established for each version as well as for the entire set of experimental items. Therefore, the same materials and test versions were used in the entailment questionnaire.

¹²That is, no disabling condition for the events described by the progressive VP was introduced by context items.

¹³See <http://www.ru.nl/celex/>

3.3. Procedure. Copies of the questionnaire were distributed to all subjects in the database meeting the following criteria: they had to be native speakers of Dutch with an age between 18 and 50 and with no history of neurological, psychiatric or cognitive disorders. The first 36 subjects who returned the questionnaire were included in the analysis.

The first page of each booklet contained the test instructions. Participants were informed that the questionnaire consisted of 160 short texts and that each comprised three sentences (the two context sentences and the critical sentence) and was followed by a pair of sentences (the probes). Subjects were instructed to read each sentence carefully and select the probe which they deemed correct on the basis of their expectations ('verwachtingen') concerning the continuation of the text ('over het vervolg van de tekst'). Subjects were asked to respond as quickly and accurately as possible and to write down a brief explanation of their answer in the blank space following the probes.

The reference to 'expectations' in the test instructions requires some explanation. Notice that we were not interested in what 'subjects semantically know' given a progressive clause, presumably that the writing activity was taking place some time in the past and that only a part of the letter was then completed. For this would amount to asking subjects *what is true at the reference time*, which is captured by the integrity constraint (9) above and requires no inference on the relevant models. Rather, we were interested in what subjects conclude about the outcomes of an action described using the progressive. That is, we wanted to know what subjects infer about goals and, more precisely, *what would be the course of events after the reference time*. In order to shift the attention of participants from the reference time to these later times, we constructed the probe pairs (P1) and (P2) using the Dutch present perfect (which focuses on the present consequences of a past event), we avoided probes of the form 'The girl has written a part of the letter', and we asked subjects to decide which of the two probes (positive or negative) matched their 'expectations' about the continuation of the narrative. Participants' written explanations provide no evidence that (P1) and (P2) were insufficient to give accurate responses such that, for instance, a probe of the form 'It is unclear whether the girl has written a letter' was necessary. Rather, subjects' comments suggest that, if the discourse implied that only a part of the letter was completed, as in (S4), then the negative probe in (P2) had to be selected.

3.4. Data analysis. Subject-based and item-based statistical analyses were carried out. For the former, we used a repeated-measures ANOVA model with Subject as the random effect, Aspectual Class (Activity or Accomplishment) and Condition Type (Neutral or Disabling) as fixed effects, and the average number of negative responses (that is, of negative probes chosen as responses) computed across items as the dependent variable. To generalize over both subjects and test items,¹⁴ we used a parallel repeated-measures ANOVA model in which Test Item (as defined above, that is, as a set of context, critical and probe sentences) was the random effect, Aspectual Class and Condition Type were the fixed effects, and the average number of negative respondents (subjects giving a negative response) computed across items was the dependent variable. Univariate *F*-tests were computed in both cases.

3.5. Results. The subject-based analysis revealed significant effects of Aspectual Class and Condition Type, and a significant interaction between the two factors (see Table 1). Neutral activities (S1) had the lowest average of negative responses ($M=2.72$, $SD=3.22$), followed by disabled activities (S2) ($M=8.06$, $SD=7.05$), neutral accomplishments (S3) ($M=10.03$, $SD=9.23$) and disabled accomplishments (S4) ($M=25.14$, $SD=8.02$). Except for neutral activities (S1), the distribution of the data in the different conditions appears rather similar, as indicated by standard deviations as well as box height and whisker length in Figure 1a. Figure 1b shows that the pattern of responses is largely as predicted by the theory and is consistent with conventional semantic wisdom: accomplishments are more sensitive than activities to the presence of a disabling condition in the discourse context. A similar pattern of effects was

¹⁴See [Clark, 1973] for the main motivation and [Wike and Church, 1976] for a critical discussion.

revealed by the item-based analysis (Table 1). Neutral activities (S1) had the lowest average of negative respondents ($M=2.35$, $SD=4.13$), followed by disabled activities (S2) ($M=7.13$, $SD=7.24$), neutral accomplishments (S3) ($M=9.54$, $SD=7.09$) and disabled accomplishments (S4) ($M=22.62$, $SD=9.04$).

	Subject-based analysis	Item-based analysis
Aspectual Class	$F(1,35)=64.763$ ***	$F(1,159)=619.240$ ***
Condition Type	$F(1,35)=112.560$ ***	$F(1,159)=237.270$ ***
Aspectual Class \times Condition Type	$F(1,35)=61.832$ ***	$F(1,159)=100.210$ ***

Significance codes: $p=0$ '****' $p=0.001$ '***' $p=0.01$ '**' $p=0.05$ '.' $p=0.1$ '' $p=1$

TABLE 1. Summary of ANOVA statistics for the entailment questionnaire.

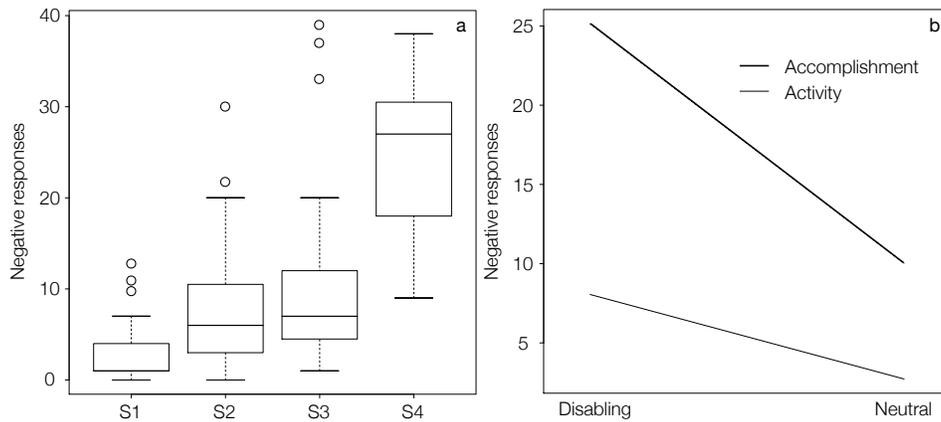


FIGURE 1. Results of the questionnaire. (a) Boxplot for the subject-based analysis. The solid line within each box represents the median, box height is equal to the interquartile range, whiskers indicate the data range and empty circles are outliers. The maximum of potential negative judgments is 40. (b) Interaction plot for the subject-based analysis. Trace endpoints represent the average number of negative responses computed across participants in each experimental condition. The maximum of potential negative judgments is 40.

4. PROCESSING MODEL

In the preceding sections we have proposed an analysis of the imperfective paradox based on the Event Calculus and we have presented some behavioral data supporting the theory. Recall that we chose to consider our semantic theory as a computational analysis in Marr's sense. We are now ready to move on to the intermediate level of analysis and consider the actual processing steps subserving the computation of discourse models.

The processing model is based on the combination of the Event Calculus and Constraint Logic Programming used in [van Lambalgen and Hamm, 2004]. The algorithms presented there specify in an abstract manner the computational steps involved in satisfying integrity constraints, that is they formalize the derivation (the proof) of a given statement within the theory. In order to derive predictions concerning the complexity and the time-course of the relevant semantic computations, it is necessary to add an explicit processing component to the formal machinery of the Event Calculus. This is the first task undertaken in this section. We shall then look at the hypotheses licensed by the resulting model.

4.1. The principle of immediacy. In a paper discussing language processing models and their neural implementation, Hagoort proposed "six general architectural principles for comprehension beyond the word level" [Hagoort, 2006]. One of these, also known as 'the principle of immediacy', is particularly relevant in this context. Immediacy has often been debated in the psycholinguistic literature [Marslen-Wilson, 1989, Jackendoff, 2002], where it is usually regarded as a hypothesis on the time-course of the access and integration of lexical meanings, and is sometimes contrasted with syntax-first models [Frazier, 1987, Friederici, 2002]. Immediacy is defined by Hagoort as

the general processing principle of unification.¹⁵ Semantic unification does not wait until all relevant syntactic information (such as word class information) is available, but starts immediately with what it derives on the basis of the bottom-up input and the left context. The corollary of immediacy is incrementality: output representations are built up from left to right in close temporal contiguity to the input signal. [Hagoort, 2006]

This general statement cannot be directly related to our Event Calculus analysis. The reason is that it refers to meaning integration only, and not to the computation of denotations and discourse models. Hagoort's formulation of immediacy is sufficient to derive processing predictions for cases of semantic composition. However, to be relevant for describing the construction of discourse models, the principle of immediacy must be thus reformulated:

Proposition 1. *The algorithms updating a minimal model so as to satisfy an integrity constraint are executed as soon as the integrity constraint is given as input.*

The present reformulation of immediacy translates into a general hypothesis concerning the time-course of the computation of discourse models: it says that minimal models are computed as soon as update instructions, in the form of integrity constraints, are given as input. Together with Theorem 1, Proposition 1 forms the core of our processing model.

4.2. The recomputation hypothesis. Let us now consider the processing steps leading to the construction of discourse models for (S3) and (S4). The first update instruction, in the form of integrity constraint (9), is introduced by the progressive VP. Since at that stage the information provided by discourse amounts to the context (C) and the main (progressive) clause of (S3) and (S4), no obstacle to the writing activity is described. Proposition 1 implies that (9) is satisfied as soon as the progressive VP is processed and Theorem 1 guarantees that a minimal model will be computed such that there is a time at which the consequent state holds. Combining immediacy with the Event Calculus suggests the hypothesis that, *when an accomplishment in the progressive is processed, a minimal model is immediately computed in which the goal state holds.*

¹⁵For the notion of 'unification' in the Event Calculus and its role in language processing see [Baggio et al., 2007].

Considering the subordinate clause, we must distinguish two processing steps. The first is carried out by adding (10) to the scenario (8), updating the discourse model according to (11), and deriving the existence of a time later than the initiation of the writing activity at which writing was terminated. The second step computes the suppression of the goal state inference by deriving the further statement that there is no time at which a complete letter is obtained. These are two distinct processing steps: for activities which are terminated by a disabling condition, like (S2), only the first step is carried out when the subordinate clause is processed; indeed, there is no goal state to be canceled, because (8c) and (8e) are not in the scenario, but still it is part of the meaning of (S2) that writing was terminated; for disabled accomplishments, like (S4), both steps must be implemented. Recall that the satisfaction of the goal state is derived in the minimal model of the progressive VP. As a consequence, the hypothesis licensed by our theory is that, *when an accomplishment in the progressive is followed by a subordinate clause describing a disabling condition, the initial minimal model is recomputed to the effect that, in the new discourse model, there is no time at which the goal state holds.*

Recomputation is thus the main processing consequence of the imperfective paradox. Predictions concerning the complexity and the time-course of semantic computations for sentences like (S1)-(S4) can be derived from our processing model. The theory predicts that the subordinate clause in (S4) involves the recomputation of the minimal model computed while processing the progressive clause, while in (S3) the initial model is simply extended (we return on the difference between ‘recomputation’ and ‘extension’ in 4.3). Furthermore, the activity cases (S1) and (S2) will also involve a monotonic extension of the initial model, such that the termination of the writing activity is computed for (S2), but no cancelation of the goal state (since a predefined goal is not involved in activities). In short, the subordinate clause in (S4) will be more complex to process compared to (S1)-(S3), as no recomputation is triggered in the latter conditions. As for time-courses, EC requires the causal and temporal information carried by the final verb ‘morste’ to activate the additional scenario clause (10), satisfy (11) and derive the non-occurrence of the consequent state. Hence, recomputation is expected to surface only at the final word in (S4).

4.3. Recomputation and perceptron learning. Below we address a few outstanding issues raised by the processing model. This will give us the opportunity to clarify our claims and to extend our framework so as to cover issues relating to the neural implementation of the theory. We will first discuss the relations between the Event Calculus, Logic Programming and connectionist networks, to then turn to the possible neurophysiological correlates of model recomputation as described above.

The reader might wonder why computing a minimal model that is incompatible with a previous model (recomputation) would be different from, and in particular more complex than computing a minimal model that monotonically extends the initial model. An answer can be found in the hypothesized behavior of the networks which compute these models. Due to the syntactic restrictions inherent in Logic Programming, the models characterized by the Event Calculus can be viewed as stable states of associated neural networks. It has been demonstrated that recurrent neural networks are sufficient and suitable for computing minimal models for propositional logic programs [Stenning and van Lambalgen, 2005, Stenning and van Lambalgen, 2007]. It has also been shown that, for a given propositional logic program, it is possible to construct a 3-layer feedforward network of binary threshold units computing the semantic operators on which the construction of minimal models is based [Hitzler et al., 2004]. The formal language underlying EC is not propositional but is a many-sorted predicate logic (see 2.1), with matters being further complicated by the use of integrity constraints. Recent research suggests however that recurrent networks can also approximate the semantic operators for first-order logic programs and their fixed points [Hitzler et al., 2004]. Moreover, as shown in [Stenning and van Lambalgen, 2007], integrity constraints can be modeled via a simple form of backpropagation called ‘perceptron learning’ [Rojas, 1996, p. 84].

In this framework, any computation on a given minimal model, such as adding a logic program clause (a scenario clause), will somehow bring the network from its initial stable state to another stable state, corresponding to the new minimal model. Nevertheless, if the neural representation proposed in [Stenning and van Lambalgen, 2007] is approximately correct, there is a vast difference in neural activity between, on the one hand, a monotonic extension of a minimal model, and a non-monotonic recomputation of a minimal model on the other. Consider first the case of a monotonic extension as envisaged by our processing model. Retrieving a clause such as (10) from semantic memory, assuming that *spill* denotes ‘spilling coffee on the *tablecloth*’, will result in the activation of a number of units (neurons) which were previously silent; but the activation state of the remaining units, including those representing the goal state (the complete letter), will remain the same. However, retrieving a different clause from semantic memory, for instance (10) where *spill* now denotes ‘spilling coffee on the *paper*’, will result in the activation of units which were silent *but also* in the readjustment of the activation patterns of units which were previously active. For instance, the neurons representing the consequent state (the complete letter) will no longer be active. This change in activation is achieved in the neural network by applications of perceptron learning. The difference between monotonic extension and non-monotonic recomputation can thus be found in the occurrence of the readjustment of connection strengths driven by perceptron learning.

These considerations suggest that non-monotonic recomputation in (S4) has more drastic consequences on neural processing as compared to the monotonic extension of a minimal model in (S3). We will now argue that a processing model based on monotonic semantics does *not* predict a similar effect. In a strictly monotonic progression of models, the goal state is necessarily not represented, since otherwise the model may have to be recomputed (which is not allowed by the monotonic logic underlying the theory). In the Event Calculus framework this means that the predicates and axioms for continuous change are not used, at most those for instantaneous change. Furthermore, because the progression of models is monotonic, one never actually computes minimal models.¹⁶ Under the assumption of monotonicity little semantic computation is going on. In particular, (S3) and (S4) will both lead to monotonic extensions of the initial model, and both extensions will be computed at the same time, that is when the final verb is processed: in (S3) the update will lead to a model in which (given the results of our entailment questionnaire) the consequent state is represented as being attained, in (S4) it will lead to a model in which the consequent state does not hold. Because there is no principled semantic or processing reason to assume that one final model would be more complex to compute than the other, it follows that a strictly monotonic semantics predicts no difference between the conditions.

The recomputation hypothesis can be tested in a dedicated neurophysiological study. Event-related potentials (derived from EEG signals) or fields (derived from MEG signals) and power changes in specific frequency bands provide direct insights into the complexity and the time-course of neural processing, as opposed to reading times and eye movements, which are indirect and cumulative measures of processing load [Luck, 2005]. An EEG study using the materials of our questionnaire as stimuli would be able to determine whether the ERP correlate of model recomputation is the N400 – the negative shift peaking around 400 ms after the onset of semantically anomalous words [Kutas and Hillyard, 1980], words with lower cloze probabilities [Kutas and Hillyard, 1984, Hagoort and Brown, 1994] and words which provide information conflicting with the discourse context [van Berkum et al., 1999, van Berkum et al., 2003] or world knowledge [Hagoort et al., 2004]. Since the computations underlying these phenomena pertain more to the domain of semantic composition than to the construction of discourse models, recomputation might evoke a different ERP effect. Preliminary results [Baggio, 2006] show that the final verb in (S4) elicits a larger sustained

¹⁶Formally, one never computes the completion of the scenario and axioms. For the notion of ‘completion’ see [van Lambalgen and Hamm, 2004].

negative shift (thus different from the transient N400s) compared to conditions (S1)-(S3). The amplitude of the ERP effect in (S4) is correlated with the number of negative responses, which is what one should observe if the sustained negativity was the brain signature of recomputation: assuming that recomputation evokes a sustained negative shift in each (S4) trial in which a negative response is given, the larger the number of negative responses (the larger the number of trials in which recomputation took place), the larger the amplitude of the ERP component revealed by averaging. Current experimental work in our EEG lab is aimed at corroborating these findings.

5. CONCLUSIONS

The work presented here is arguably far from a proper integration of theoretical semantics, psycholinguistics and the cognitive neuroscience of language. Nevertheless, it is a concrete step in that direction. Furthermore, our approach can be extended to a number of related phenomena. The first is aspectual coercion, a subtractive form of which is involved in the activity ‘writing letters’ compared to the accomplishment ‘write a letter’. The computations which may result in specific ERP effects are those involved in removing the scenario clauses regimenting the behavior of the goal state [van Lambalgen and Hamm, 2004, pp. 172-173]. The second issue is non-veridical uses of ‘before’, which show entailment patterns that are similar to those of accomplishments in the progressive [Cresswell, 1977, Åqvist et al., 1978]. An analysis of temporal connectives in the Event Calculus might yield new insights into the processing of these constructions. Lastly, the recomputation hypothesis can be applied to several non strictly semantic phenomena that have been hypothesized to involve non-monotonic inference, such as presuppositional denials and implicature cancellations in pragmatics and the suppression effect with conditionals in the psychology of reasoning [Stenning and van Lambalgen, 2007].

Acknowledgments. This paper benefited from many useful discussions with Travis Choma, Bart Geurts, Peter Hagoort and Karl Magnus Petersson. Thanks also to Ira Noveck and to two anonymous reviewers for their helpful comments. We are grateful to the Netherlands Organization for Scientific Research for support under grant NWOCOG/04-19.

REFERENCES

- [Åqvist et al., 1978] Åqvist, L., Guenther, F., and Rohrer, C. (1978). Definability in ITL of some subordinate temporal conjunctions in english. In Guenther, F. and Rohrer, C., editors, *Studies in Formal Semantics*, pages 201–221. North-Holland.
- [Baggio, 2006] Baggio, G. (2006). Computing and recomputing discourse models. Poster presented at the 12th Annual Conference on Architectures and Mechanisms for Language Processing, August 30th - September 1st 2006, Nijmegen.
- [Baggio et al., 2007] Baggio, G., van Lambalgen, M., and Hagoort, P. (2007). Language, linguistics and cognition. In Stokhof, M. and Groenendijk, J., editors, *Handbook of Philosophy of Linguistics*. Elsevier.
- [Clark, 1973] Clark, H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, 12(4):335–359.
- [Cook et al., 2001] Cook, A., Limber, J., and O’Brien, E. (2001). Situation-based context and the availability of predictive inference. *Journal of Memory and Language*, 44:220–234.
- [Cresswell, 1977] Cresswell, M. (1977). Interval semantics and logical words. In Rohrer, C., editor, *On the Logical Aspects of Tense and Aspect*, pages 7–29. Narr.
- [de Swart, 1998] de Swart, H. (1998). Aspect shift and coercion. *Natural Language and Linguistic Theory*, 16:347–385.
- [Dowty, 1979] Dowty, D. (1979). *Word meaning and Montague grammar*. Reidel, Dordrecht.
- [Frank et al., 2003] Frank, S., Koppen, M., Noordman, L., and Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, 27:875–910.
- [Frazier, 1987] Frazier, L. (1987). Sentence processing: A tutorial review. In Coltheart, M., editor, *Attention and Performance*, volume XII. Erlbaum Press.
- [Friederici, 2002] Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6:78–84.
- [Geurts and van der Slik, 2005] Geurts, B. and van der Slik, F. (2005). Monotonicity and processing load. *Journal of Semantics*, 22:97–117.
- [Hagoort, 2006] Hagoort, P. (2006). The Memory, Unification, and Control (MUC) model of language. In A.S. Meyer, L.R. Wheeldon, A. K., editor, *Automaticity and Control in Language Processing*. Psychology Press.
- [Hagoort and Brown, 1994] Hagoort, P. and Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In Clifton, C., Frazier, L., and Rayner, K., editors, *Perspectives on Sentence Processing*, pages 45–81. Lawrence Erlbaum Associates.

- [Hagoort et al., 2004] Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304:438–41.
- [Hitzler et al., 2004] Hitzler, P., Hölldobler, S., and Seda, A. (2004). Logic programs and connectionist networks. *Journal of Applied Logic*, 3(2):245–272.
- [Jackendoff, 2002] Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- [Kutas and Hillyard, 1980] Kutas, M. and Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207:203–205.
- [Kutas and Hillyard, 1984] Kutas, M. and Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163.
- [Landman, 1992] Landman, F. (1992). The progressive. *Natural Language Semantics*, 1(1):1–32.
- [Luck, 2005] Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press.
- [Marr, 1982] Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman.
- [Marr and Poggio, 1976] Marr, D. and Poggio, T. (1976). From understanding computation to understanding neural circuitry. *MIT AI Memos*.
- [Marslen-Wilson, 1989] Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In Marslen-Wilson, W., editor, *Lexical representation and process*. The MIT Press.
- [McMillan et al., 2005] McMillan, C., Clark, R., Moore, P., Devita, C., and Grossman, M. (2005). Neural basis of generalized quantifier comprehension. *Neuropsychologia*, 43:1729–1737.
- [Michaelis, 2001] Michaelis, L. (2001). A unification-based model of aspectual type-shifting. University of Colorado at Boulder; see <http://spot.colorado.edu/~michaeli/>.
- [Minsky, 1977] Minsky, M. (1977). Frame-system theory. In Johnson-Laird, P. and Watson, P., editors, *Thinking*, pages 355–376. Cambridge University Press.
- [Parsons, 1990] Parsons, T. (1990). *Events in the Semantics of English*. MIT Press.
- [Reichenbach, 1947] Reichenbach, H. (1947). *Elements of Symbolic Logic*. Macmillan.
- [Rojas, 1996] Rojas, R. (1996). *Neural Networks: A Systematic Introduction*. Springer-Verlag.
- [Schank and Abelson, 1977] Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates.
- [Stenning and van Lambalgen, 2005] Stenning, K. and van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, 29(6):919–960.
- [Stenning and van Lambalgen, 2007] Stenning, K. and van Lambalgen, M. (2007). *Human Reasoning and Cognitive Science*. MIT Press.
- [van Berkum et al., 1999] van Berkum, J., Hagoort, P., and Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11:657–671.
- [van Berkum et al., 2003] van Berkum, J., Zwitterlood, P., Hagoort, P., and Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17:701–718.
- [van Dijk, 1980] van Dijk, T. (1980). *Macrostructures*. Lawrence Erlbaum Associates.
- [van Lambalgen and Hamm, 2004] van Lambalgen, M. and Hamm, F. (2004). *The Proper Treatment of Events*. Blackwell.
- [Vendler, 1957] Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 66(2):143–160.
- [Wike and Church, 1976] Wike, E. and Church, J. (1976). Comments on clark’s “The language-as-fixed-effect fallacy”. *Journal of Verbal Learning and Verbal Behaviour*, 14(3):249–255.

DO LANGUAGE PROCESSING DATA SUPPORT COMPOSITIONALITY?

GIOSUÈ BAGGIO, MICHEL VAN LAMBALGEN, AND PETER HAGOORT

GIOSUÈ BAGGIO

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen

MICHEL VAN LAMBALGEN

Institute for Logic, Language and Computation

University of Amsterdam

PETER HAGOORT

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen *and*

Max Planck Institute for Psycholinguistics, Nijmegen

Address correspondence to:

GIOSUÈ BAGGIO

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen

P.O. Box 9101

NL-6500 HB Nijmegen

The Netherlands

Phone: +31.024.3610980

Fax: +31.024.3610989

E-mail: g.baggio@fcdonders.ru.nl

Do not quote -- To appear in The Oxford Handbook of Compositionality

1. INTRODUCTION

The title of this chapter invites another, more fundamental question: *can* experimental data support, or for that matter undermine, compositionality? After all, we are often reminded, compositionality is formally vacuous: any grammar can be given a compositional semantics (Janssen 1986; Zadrozny 1994),¹ which implies compositionality is also *empirically* vacuous: if a compositional treatment of any linguistic phenomenon can be given, compositionality is always upheld by the data. More precisely, the meaning of any complex expression can be viewed as a function of the meanings of its constituent parts and the syntactic mode of combination, *provided enough complexity is built into the structures involved*, that is the lexicon and the syntax. These are not motivated on independent grounds, as their characterization serves the sole purpose of yielding a compositional theory (Groenendijk & Stokhof 1991).

The need for an independent motivation of theories of syntax and lexical semantics is precisely the issue here. Our aim is to show that, even though there often is a way to salvage compositionality in the face of empirical data, the choices one has to make in order to do so have consequences which may be untenable given the cognitive and neural constraints on language comprehension, production and acquisition. This in turn may give us some hints as to how to answer the question posed by the title. But let us start with the most basic of questions: *why* compositionality? We will now give a sketch of the main arguments, which will be refined in the course of the discussion.

1.1. **The productivity argument.** A familiar argument in favor of compositionality starts from a perceived tension between the infinity of language and the finiteness of the brain. There are infinitely many sentences in any natural language, but the brain has only finite storage capacity, and it therefore falls to syntax to provide a finitely describable procedure for generating an infinite class of sentences. Furthermore, so the argument goes, a speaker of any language is able to understand a sentence she has never heard before, or to express a meaning she has never expressed before, and in that sense she knows the infinitely many different meanings of the infinitely many sentences of that language. Therefore, semantics is also under the obligation to come up with a finitely describable engine that generates all possible sentence meanings for the given language (Katz & Fodor 1963).

¹See Kazmi & Pelletier (1998) and Westerståhl (1998) among others for a critical discussion.

30 Compositionality provides a seemingly efficient way to satisfy these desiderata. There are only finitely many words in the lexicon and syntax can have only finitely many rules of combination. Here compositionality comes into play:

PRINCIPLE OF COMPOSITIONALITY *The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.*

35 If meanings are generated in such a way that compositionality is satisfied, then it seems that all possible sentence meanings can be finitely generated. Now, although compositionality is a guiding principle of *formal* semantics, the standard motivation as sketched above partly appeals to the *psychological* notions of comprehension and production, while at the same time invoking the patently non-psychological infinity of language. A quick way to dismiss
40 the argument from productivity is therefore to deny that language is infinite, even in the sense of *potential* infinity. A moment's reflection shows however that the issue is not really about infinity: substituting a large finite number for infinity does not change the essence of the productivity argument, which is that not every sentence that can be understood or produced given human cognitive limitations is stored. So, while there is no reason to have
45 a semantic theory that explains comprehension of nested center embeddings of arbitrary depth, it is also not the case that all sentences with center embeddings with depth, say, ≤ 6 can be stored. In other words, psychologically speaking, the real issue is about 'the balance between storage and computation', and the role compositionality plays there. And it might seem that compositionality always leads to the most efficient architecture in this respect.

50 That this is not necessarily so can be illustrated using an example from Keenan (1979). In an adjective–noun construction, the noun is the argument fed into the adjective, which is viewed as a function. Keenan observes that the interpretation of the function word seems to be determined by its argument: compare for instance the different meanings of the adjective 'flat' in 'flat tyre', 'flat beer', 'flat note' etc. It is of course technically possible, as Keenan
55 notes, to replace the single function 'flat' by a disjunctively defined function, where each of the disjuncts corresponds to a separate meaning for 'flat', with suitable selection restrictions on the argument. However, this technical solution is surely paradoxical: compositionality was invoked to account for productivity, which seemed hard to explain in terms of storage only; but, in this case, compositionality can apparently be salvaged only by *increasing* the

60 demand on storage! From a processing perspective, it would be much better if there were a single computational mechanism generating the meaning of a *flat* + *noun* construction, starting from a single basic meaning of ‘flat’. These considerations show that the principle of compositionality is affected by its ambiguous status: as a formal desideratum on the one hand, and a processing hypothesis on the other. Let us now turn to the former aspect. The remainder of this chapter will focus on compositionality as a processing principle.

1.2. **The methodological argument.** A second argument is that compositionality is needed as a constraint on doing semantics, as an essential part of the explanatory enterprise (Janssen 1997; Dever 1999). For instance, if one has to explain why in the ‘donkey sentence’

(1) If a farmer owns a donkey, he beats it.

70 the DP ‘a donkey’ has universal force, it will not do to say: ‘well, in this context it simply has universal force’. An account that starts out with the existential interpretation of the DP, and then shows how its being embedded in the antecedent of the conditional changes its interpretation from existential to universal, has at least the appearance of an explanation.

The trouble with the methodological argument is that compositionality is highly theory 75 dependent (Partee et al. 1990). Ideally, when looking for an explanation of a given linguistic phenomenon, one takes the syntax and the semantics to be fully specified formal systems. It is then a definite question whether that phenomenon allows for a compositional treatment. If it doesn’t, one may take this as a cue for changing the semantics. In practice, however, the explanation of a new phenomenon of interest often leads to changes in *both* syntax and 80 semantics. Compositionality then becomes a soft constraint indeed.

It seems to us the much-needed methodological constraints have to be sought elsewhere, in a tighter regimentation of syntactic and semantic theories. From our perspective, these constraints should be cognitively motivated, in the sense that formal theories of syntax and semantics should be viewed as ‘computational level theories’ (Marr 1982) of actual syntactic 85 and semantic mechanisms (Baggio & van Lambalgen 2007; Baggio et al. 2008b). In the ideal case, it then becomes an empirical question whether syntax and semantics communicate as compositionality says they do. This leads us to the third argument, in which the status of compositionality as a processing principle becomes more prominent.

1.3. **The modularity argument.** A third argument leading to compositionality is suggested
90 by a view of the language faculty as a ‘cognitive module’. Fodor (1983) lists nine properties
characterizing modular systems: domain specificity, fast and mandatory operation, limited
central access to modular representations, informational encapsulation, shallow outputs,
fixed neural architecture, specific breakdown patterns, and characteristic ontogenic pace and
sequencing. Of these, the most relevant for our purposes is *informational encapsulation*. This
95 is the idea that perceptual systems – language included – are relatively impenetrable to the
bulk of the knowledge internally represented by the organism. Informational encapsulation
says that there are tight constraints on the flow and handling of extra-modular information
within the module *prior to* the production of an output. Interactions between the module’s
operations and its environment can therefore occur only at the level of central systems.

100 To a certain extent, informational encapsulation is assumed by any cognitive model of
language – which is not to say that all component-based architectures (Jackendoff 1997) are
modular in Fodor’s sense. Fodor’s (and Chomsky’s) original view of modularity was that a
grammar’s generative power can be captured by a *single* module, which comprises a finite
repository of lexical meanings and a finite repertoire of syntactic rules.² Rules for *semantic*
105 computation (inference is a paradigmatic case) fall within the province of central systems.
It is fairly easy to see that the computations performed by this kind of modular machine are
those regimented by compositionality: the output produced by the module (the meaning of
a complex expression) is a function of the knowledge available to the module (the lexicon
and the syntax). But this is not the only modular architecture supporting compositionality.
110 For instance, one could postulate *two* modules: a module which produces syntactic analyses
of clauses, which are then fed into another module containing meanings for the lexical items
and combination procedures corresponding to syntactic operations, and outputs a semantic
representation of the clause. Compositionality would then constrain the kind of traffic that
can occur between the two modules.

115 So regardless of one’s choice of modular architecture, compositionality remains relevant
insofar as it acts as a counterpart of information encapsulation at the level of the description
of linguistic structure. The link between compositionality and informational encapsulation

²The module also contains mechanisms for phonological decoding, but let us ignore these for simplicity.

can be made more explicit: if the composition of meanings is not affected by extra-modular knowledge, then one can characterize the meaning of any complex expression as a function
 120 of the meanings of its constituent parts and syntactic rules, all of which are readily available within the module.³ Clearly, this hinges very much on what one assumes is contained in the module(s) – and this will be a recurrent theme in this chapter. What bears some emphasis here is that the degree to which a system is informationally encapsulated can be determined only based on empirical data. Hence, with this ‘argument from modularity’ in place, it also
 125 becomes possible to treat compositionality as a processing principle, that is, as a constraint on the kind of structures that can be computed during language processing.

2. COMPOSITIONALITY AS A PROCESSING PRINCIPLE

2.1. **A first approximation.** The issue which we set ourselves to address is how to constrain and refine compositionality based on experimental data and cognitive considerations. One
 130 could start from the observation that ‘function’ in the definition of compositionality needs to refer to some computable process, and that inputs – lexical meanings and syntactic rules or constraints – must be given incrementally:

INCREMENTAL COMPOSITION *The meaning of a complex expression at some processing stage σ is computed based on the constituent expressions processed at
 135 σ and of the syntactic structure built up at σ .*⁴

This definition is silent as to whether meaning assembly involves the lexicon and the syntax *only*, or whether other sources of information can enter the composition process. For this we need another definition, which we choose to formulate in terms of constraints:

SIMPLE COMPOSITION (i) *The meanings of elementary expressions are the only
 140 constraints on content in the computation of the meaning of a complex expression.*
 (ii) *The syntax of elementary expressions is the only constraint on structure in the computation of the meaning of a complex expression.*

³Jackendoff (1997) makes the same point: “The hypothesis of syntactically transparent semantic composition has the virtue of theoretical elegance and constraint. Its effect is to enable researchers to isolate the language capacity – including its contribution to semantics – from the rest of the mind, as befits a modular conception.”

⁴For simplicity, we may assume that each word corresponds to a processing stage, and vice versa. An important theoretical question is whether assuming finer-grained processing steps would lead to local inconsistencies between incrementality and compositionality.

The notion of ‘weak’ or ‘enriched composition’ (Pustejovsky 1995; Jackendoff 1997) follows from allowing further constraints on content and structure. The distinction between content
 145 and structure is admittedly vague, and can only be made clearer based on particular formal theories of syntax and semantics – recall compositionality’s theory-dependence (Partee et al. 1990). We could go even further and observe that, at least in formal approaches to grammar, the distinction vanishes, and one is left with a purely syntactic analysis of meaning in some logical language. That said, it remains a distinction worth keeping, especially because the
 150 brain appears to honor it, as is reflected by the different electrophysiological traces left by morphosyntactic and semantic manipulations (see 3.3 below). Let us now see how two of the arguments for compositionality presented above fare in light of the new definitions.

2.2. Productivity and modularity reconsidered. Compositionality’s plausibility is usually argued for by means of a rhetorical question, such as ‘what other mechanism could explain
 155 the productivity of language?’, as if posing the question would already dictate the answer. To address this point, it pays to be more precise regarding the exact technical implications of simple composition. Consider two supposed consequences, due to Hintikka (1983):

CONTEXT INDEPENDENCE THESIS *The meaning of an expression should not depend on the context in which it occurs.*

160 INSIDE-OUT PRINCIPLE *The proper direction of a semantic analysis is from the inside out.*⁵

Together, these consequences suggest that we ought to take the notion of a ‘function’ in the formulation of the principle very seriously. Semantic computation of complex expressions is function application, so *the meanings of simple expressions are not changed by virtue of the fact*
 165 *that they occur as arguments of the function.*⁶ This shows that the principle of compositionality is tied to one very particular form of linguistic productivity, exemplified by the usual rules: if S is a sentence, one can form a new sentence ‘I think that S ’; if S_1, S_2 are sentences, one can form ‘ S_1 and S_2 ’; etc. But there are other forms of productivity in natural languages,

⁵‘Inside-out’ is here synonymous with ‘bottom-up’, that is, from elementary meanings to complex meanings.

⁶Just as the number 2 has the same meaning, whether it occurs as argument in the function 3^x or in the function $3 + x$. This is obvious in mathematics, but not in natural language, as we shall see.

which do not have this function-like character. An example is the progressive construction
 170 in English when applied to stative verbs (Croft 1998, p. 70):

- (2) a. She resembles her mother.
 b. *She is resembling her mother.
 c. She is resembling her mother more and more every day.

'Resemble' is a stative verb, and this seems to be clinched by (2-b), which clearly shows that
 175 the progressive is not applicable. Yet, in a suitable context the progressive *is* applicable, as
 in (2-c), where it imposes a dynamic meaning upon 'resemble': resemblance now comes in
 degrees that can vary over time. Therefore, 'resemble' changes its meaning depending on
 the context in which it occurs, contradicting context independence. This change of meaning
 can still be predicted once one assumes that the progressive construction has a meaning of
 180 its own, which it imposes upon that of the verb. This imposition of meaning is moreover
 productive in that it applies to many stative verbs. Simple composition as made precise
 by Hintikka can explain this particular form of productivity only by assuming multiple
 meanings of stative verbs, where the progressive selects for a dynamic meaning (recall the
 analysis of 'flat' in 1.1).

185 While compositionality can be salvaged formally, experiments on language processing
 might rule out such *ad hoc* manoeuvres. There is a very different computational account of
 what goes on in comprehending sentences such as (2-c) that emphasises the recomputation
 of the meaning of 'resemble' which takes place when the adverbial phrase 'more and more
 every day' is processed (van Lambalgen & Hamm 2004, Chapter 11). The two accounts thus
 190 differ in their processing consequences: simple composition in the Hintikka sense leads to
 search and selection of the appropriate meaning but not to recomputation, as in the second
 account. These operations might give rise to different neurophysiological responses (Baggio
 et al. 2008a), so in principle the two accounts can be tested experimentally.

If recomputation were to be supported, where would that leave compositionality and the
 195 work it is supposed to do? The strict reading as embodied in Hintikka's principles presents
 semantic composition as entirely inside-out/bottom-up. The recomputation account is also
 partially outside-in/top-down. In theory, this has the consequence that the meaning which

is assigned to an expression is always provisional, a situation that cannot occur on a literal reading of compositionality (but see 3.1). However, there is room for both accounts, because
 200 the productivity of language is a two-dimensional phenomenon. On the vertical dimension, there is productivity due to increased syntactic complexity; here simple composition has an important role to play. There is however also a horizontal dimension to productivity – here it is not so much syntactic complexity that increases, but new meanings of a given clause are produced by varying syntactic or semantic aspects of other clauses. Thus, if we replace
 205 the adverbial ‘more and more every day’ in (2-c) with ‘in broad daylight’, the progressive is no longer felicitous, because the verb ‘resemble’ then has its default stative meaning. The horizontal dimension of productivity seems to call for some form of enriched composition, that is, ways of combining meaning that allow top-down influences and recomputation.

Two forms of top-down computation seem consistent with informational encapsulation,
 210 and thereby with its corollary compositionality. First, the information fed back comes from and remains within the module, as when it is stored in the lexicon. Second, information is fed back into the module from another module or from a central system *after* the production of an output. This is a rather trivial way to preserve informational encapsulation in the face of top-down computation. The former can be dismissed using an argument due to Hodges
 215 (2006), which is based on what we might call ‘top-down composition’. It takes its cue from the following principle (Frege 1884):

CONTEXT PRINCIPLE *Elementary expressions do not have meaning in isolation, but only in the context of (as constituents of) complex expressions.*

In briefest outline, Hodges’ proposal is this. The syntax is defined in terms of constituent
 220 structure in such a way that, if e is an expression occurring in a sentence S , then S can be viewed as $G(e)$, where $G(x)$ is a syntactic frame with open slot x . Now define an equivalence relation \sim on expressions by putting

$$e \sim f \text{ iff for all } G(x): G(e) \text{ is a sentence iff } G(f) \text{ is, and } G(e) \text{ is acceptable in the same contexts as } G(f).$$

225 The *Fregean value* of e is the equivalence class of e modulo \sim . Hodges shows that taking the Fregean value of e as its meaning yields a compositional semantics. Therefore, if we assume that the module contains all Fregean values, modularity is restored. Moreover, this notion

of meaning is pleasingly context-sensitive. Hodges gives the example of sentences which have different social connotations such as ‘he is intoxicated’ vs. ‘he is pissed’ (Hodges 2006, p. 12). The contexts in which these sentences are acceptable are very different. In this sense, these two sentences have different meanings in Hodges’ semantics, whereas they would be treated as synonymous with ‘he is drunk’ in a more standard framework.

Although Hodges’ proposal formally restores compositionality, it does so in a way that upsets the balance between storage and computation, and renders it unclear how meanings can be acquired. The Fregean value of e is defined by means of all possible uses of e , and it is doubtful that these are available to a young language learner. The point can be amplified by considering what it means to *know* the Fregean value of e . There are two components to this:

- (i) one must know which f are equivalent to e , which means that for all $G(x)$ such that $G(e)$ is a sentence iff $G(f)$ is, $G(f)$ is acceptable in exactly the same contexts as $G(e)$ – this requires one to know for all sentences $G(e)$ and all contexts C , whether $G(e)$ is acceptable in C ;
- (ii) one must know which f are not equivalent to e , which means that for all G such that $G(e)$ and $G(f)$ are sentences, one must know a context in which $G(e)$ is acceptable and $G(f)$ is not, or vice versa.

Natural (as opposed to formal) languages may not incorporate the concise representations generating the knowledge required. This implies that the storage component must already contain a great deal of information about the sentences that can be constructed from e , and the contexts in which these are acceptable. Intuitively, this goes against the purpose of a modular architecture, and it also goes against the original motivation for compositionality as easing the burden on storage – recall the argument from productivity. We therefore tend to read Hodges’ result, when applied to natural languages, as showing the implausibility of an architecture in which context sensitivity is achieved by storing extra information within the module, rather than by relaxing informational encapsulation to allow cross-module talk. In brief, the *a priori* arguments considered here show that simple composition is not enough to account for the full range of factors which make language productive. Let us now ask if similar considerations are suggested by experimental work bearing on compositionality.

3. EXPERIMENTAL DATA BEARING ON COMPOSITIONALITY

3.1. **Semantic illusions.** One assumption that underlies many semantic theories is that full
260 lexical meanings are used in the composition process. In most formal semantics, this choice
is forced by the ontology. A lexical meaning is just a placeholder for a typed entity, and this
accounts for why inputs cannot be partial objects. Moreover, meaning assembly amounts to
type composition, which explains why the inputs are still recognizable in the end product.
In lexical semantics, by contrast, lexical meanings are complex, yet compact representations,
265 such as algorithms (Miller & Johnson-Laird 1976), conceptual structures (Jackendoff 1983)
etc. One possible refinement of compositionality is to allow partial representations to be
recruited during processing.

Relevant to this issue is the well-known ‘Moses illusion’ (Erickson & Matteson 1981).
When asked ‘How many animals of each sort did Moses put on the ark?’, subjects tend to
270 respond ‘two’ without questioning the (false) presupposition that Moses was the biblical
character who did that. Similar results have been obtained with questions such as ‘After
an aircrash, where should the survivors be buried?’ (Barton & Sanford 1993) or ‘Can a man
marry his widow’s sister?’ (Sanford 2002). Hearers seem to be processing these questions
superficially enough to miss the false presuppositions. Ferreira et al. (2002) and Ferreira
275 & Patson (2007) suggest that these data challenge compositionality. Consider the Moses
question. It would seem that the meaning computed by hearers and the meaning derived
compositionally are in some important respect different. If the former were a function of the
meanings of the constituents and the syntax, ‘Moses’ would mean Moses and hearers would
notice the false presupposition. Here is what seems a clear instance of a non-compositional
280 process.

One might argue against this conclusion by emphasising that the data just show that
“each word in a sentence does not contribute its full meaning” (Sanford & Sturt 2002). The
‘full meaning’ of ‘Moses’ need not be retrieved, only some feature made salient by the
context (Sanford & Sturt 2002; Sanford 2002). This may be ‘biblical character’, ‘patriarch’
285 or some other feature responsible for the semantic proximity of ‘Noah’ and ‘Moses’ (van
Oostendorp & de Mul 1990). Feature sharing is what gives rise to the illusion. Nonetheless,
the fact that the lexicon may be a locus of shallow processing (or retrieval, as in this case)

does not speak against compositionality. Simple composition entails that the lexicon is the only provider of content for complex meanings, but not that *full* lexical representations must
 290 be used. The latter seems too strong a requirement to press upon either simple or enriched composition. For if there are such entities as ‘full lexical meanings’ – and there are reasons to be skeptical about that – they can hardly be used on most processing tasks, due to the massive amount of information, presumably continuous with non-lexical knowledge, that would be fed into the composition process. Incremental, context-sensitive feature selection
 295 during the retrieval or word meaning would thus be the default. Semantic illusions would become a special case, in which some critical semantic feature is shared between the word triggering a true presupposition (‘Noah’) and the word triggering a false one (‘Moses’). Compositionality can therefore be refined to accommodate semantic illusions, by allowing composition to make use of partial lexico-semantic representations.

300 **3.2. Misinterpretations.** Ferreira and colleagues investigated cases of misinterpretation in language processing. Some of these involve garden-path sentences:

(3) While Anna dressed the baby played in the crib.

The ‘garden-path model’ (Frazier 1987) hypothesises that ‘the baby’ is initially parsed and interpreted as the direct object of ‘dressed’. Only when the verb ‘played’ is encountered the
 305 syntactic and semantic representations are revised to the effect that ‘the baby’ is the subject of ‘played’. One question here is whether the initial representation, featuring ‘the baby’ as a direct object, is at all maintained in memory. Christianson et al. (2001) show that, while readers correctly respond for the affirmative to ‘Did the baby play in the crib?’, they also give a positive answer to ‘Did Anna dress the baby?’. No grammar or parser on the market
 310 would allow for the same token NP to play two functional roles, subject and direct object. And yet, this appears to be precisely the interpretation computed by readers. Ferreira et al. (2002) take this as “clear evidence that the meaning people obtain for a sentence is often not a reflection of its true content” – that is, it is not built up compositionally.

Does the existence of ‘lingering misinterpretations’ demonstrate that the processing of
 315 garden-path sentences is non-compositional, as suggested by Ferreira and colleagues? There are at least two ways of accounting for the data, from which different answers ensue. On

one account, the last interpretation subjects come up with is that while Anna dressed, the baby played in the crib, which corresponds to the revised parse whereby ‘the baby’ is the subject of ‘played’, and is no longer the direct object of ‘dressed’. Interpretation is therefore non-monotonic, that is, allowing for revisions on earlier structures. This however requires a refinement of compositionality which we already introduced – incremental composition. On this view, both the initial and the revised interpretations can be derived compositionally, and simple composition seems enough in this case. The persisting misinterpretation would be rather an effect of memory architecture or neural implementation. One aspect of the data of Christianson et al. (2001) which supports this view is that misinterpretations are more frequent when the head of the misanalyzed phrase occurs early. That is, misinterpretations are more likely to persist the longer they have been part of the initial discourse model.

On the second account, the final interpretation is that Anna dressed the baby while the baby played in the crib. This meaning can hardly be derived compositionally. First, because there is only one token of ‘the baby’ among the constituents of the sentence, while the final interpretation we have assumed involves two occurrences of it: one as the recipient of the dressing action, the other as the agent of the playing activity – hence the ‘constituent parts’ aspect of the definition of compositionality is out. Second, because syntax does not allow a phrase to play two distinct functional roles simultaneously – hence the ‘syntax’ part of the definition is out. To derive the meaning above, one needs a mechanism that copies the token ‘the baby’ and makes both instances available for interpretation. Such a mechanism does make processing non-compositional (Ferreira et al. 2002; Ferreira & Patson 2007). In brief, misinterpretations of garden-path sentences challenge compositionality, unless we assume that early semantic material lingers in memory also during later stages, but is not part of the discourse model computed on the basis of the revised syntactic analysis.

3.3. Interlude: event-related brain potentials (ERPs). One aspect of enriched composition is that the stored meanings of elementary expressions are not the only source of content for complex meanings. It can be suggested that this challenges informational encapsulation, if one can demonstrate that such additional semantic information is handled by the system (or module, component etc.) before an output is produced. Some experimental techniques

in cognitive neuroscience allow one to make timing inferences, the most sensitive and direct of which is based on event-related brain potentials (ERPs).

The post-synaptic currents generated by the brain (called ‘electroencephalography’, EEG) can be recorded by placing a number of electrodes on the scalp, amplifying the signal and plotting the observed voltage changes as a function of time (Luck 2005). ERPs are defined as electrical potentials time-locked to some event of interest, such as the onset of a stimulus. Averaging over a relatively large number of trials of the same condition is the most widely used approach to obtaining ERPs. The assumption is that noise, defined as the brain activity which is not evoked by the cognitive process of interest, is randomly distributed in terms of polarity, latency (with respect to the onset of the stimulus or the response) and amplitude in the EEG trace. As a consequence, averaging will progressively reduce noise and reveal the ERP signals. ERPs seem particularly suited for the study of information processing as they provide an excellent window on the temporal dynamics of neural activity (Donchin 1979; Van Berkum 2004). Furthermore, ERPs have a clear advantage over cumulative measures like eye movements, reading and reaction times in that they provide *qualitative* information about which processing stages are affected by the experimental manipulation (Luck 2005). For example, manipulations of linguistic complexity and acceptability at different levels of linguistic structure result in modulations of different ERP components (Kutas & Hillyard 1980; Osterhout & Holcomb 1992; Hagoort et al. 1993), as do manipulations taxing memory systems involved in language processing (Mecklinger et al. 1995; Müller et al. 1997).

An ERP component that is particularly relevant here is the N400. This is a negative shift starting around 250 ms after word onset, peaking at 400 ms and lasting until approximately 550 ms. Every content word elicits an N400, but the amplitude of the component, relative to a control condition, is dependent upon the degree of semantic relatedness of the given word with its context, be it a sentence (Kutas & Hillyard 1980; Kutas & Hillyard 1984; Hagoort & Brown 1994) or a discourse (van Berkum et al. 1999; van Berkum et al. 2003). There is some evidence that the N400 does not reflect lexical access, but the *integration* of a word’s meaning into the unfolding semantic representation (Brown & Hagoort 1993). In the light of incremental composition, the N400 can thus be seen as an index of the complexity of

375 combining the meaning of the given word with the meanings of the expressions already
processed.

Another ERP component that will be of interest here is the P600. This is a positive shift starting around 500 ms following the onset of the word and lasting for about 500 ms. The component was called P600 (Osterhout & Holcomb 1992) after its polarity and peak latency
380 and SPS (Syntactic Positive Shift) on the basis of its functional properties (Hagoort, Brown, & Groothusen 1993). Larger P600 effects are elicited by sentences containing violations of syntactic constraints, such as phrase structure, subcategorization, agreement (case, gender, number), temporarily syntactically ambiguous sentences, garden-path sentences, as well as constructions with increased syntactic complexity.⁷ In relation to incremental composition,
385 we can take the P600 as an index of the time and resources involved in attaching the given word in the syntactic representation computed thus far (Hagoort 2003). How do N400 and P600 data bear on compositionality?

3.4. **World knowledge.** Relevant to this question is an ERP study by Hagoort et al. (2004), using true (4-a), false (4-b) and semantically anomalous (4-c) sentences:⁸

- 390 (4) a. Dutch trains are yellow and very crowded.
b. Dutch trains are white and very crowded.
c. Dutch trains are sour and very crowded.

The words ‘white’ and ‘sour’ evoked very similar N400s, in both cases larger than the N400 elicited by ‘yellow’. Integrating the meanings of ‘white’ and ‘sour’ in the ongoing semantic
395 representation is thus relatively hard. This suggests that, upon encountering ‘Dutch trains’, features are retrieved which code for the colour of Dutch trains – yellow, blue and crimson, not white – and are responsible for the additional processing costs associated with ‘white’.

While it is notoriously hard to define ‘core’ semantic features, separating linguistic from world knowledge, it is nevertheless possible to identify features which are *invariant* across
400 the individuals and communities using the relevant word. That ‘sour’ cannot be applied to

⁷See Hagoort, Brown, & Osterhout (2001), Hagoort (2003) and Friederici (2002) and for discussion and different interpretations of the P600.

⁸The stimuli were in Dutch and participants were native Dutch speakers.

(Dutch) trains seems a piece of invariant knowledge, and in that sense is a fact of linguistic predication ('linguistic knowledge'). However, trains differ in colour and other properties in space and time, hence that 'white' cannot be applied to Dutch trains reflects a contingent state of affairs which not all users of the expressions 'train' or 'Dutch train' may be aware of
405 ('world knowledge'). The N400 data of Hagoort et al. show that 'white' is hard to integrate in the sentence context, which in turn suggests there is something in the meaning of 'Dutch trains' which makes integration hard. This must be knowledge that Dutch trains are yellow, blue etc., and not white. It thus seems that during processing meanings are computed – for instance, of the compound 'Dutch trains' – encompassing invariant *and* community-specific
410 semantic information, that is, linguistic *and* world knowledge. As for compositionality, this may mean two things, depending on one's view of the lexicon: either the lexicon includes declarative memory in its entirety, and then simple composition seems enough to account for the similarity between the N400 effects, or the lexicon includes invariant meanings only, and then enriched composition – the thesis that the lexicon is not the only source of semantic
415 content – is necessary to explain the observed N400s.⁹

3.5. Co-speech gestures. At least another ERP study reporting modulations of the N400 seems relevant for our discussion. Özyürek et al. (2007) show that larger N400s are elicited by co-speech gestures which do not match with the semantic content of the accompanying sentence. This demonstrates that semantic information from different modalities – speech
420 and gesture, in this case – is integrated in the same time frame. The choice mentioned above between two views of the lexicon applies here too. If schematic representations of gestures are stored in declarative memory, which is assumed to be fully contained in the lexicon, simple composition seems enough to explain the data. If however gesture schemes are not part of the lexicon, some form of enriched composition must occur, whereby the semantics
425 of elementary expressions is just one source of content for complex meanings. This choice between two views of the lexicon, from which the status of compositionality in many cases depends, shows how severe is the problem of compositionality's theory-dependence, and how pressing is the need for realistic constraints on the components of the grammar.

⁹For other results bearing on the issue of world knowledge and semantic composition, see Münte et al. (1998), Ferretti et al. (2007) and Baggio et al. (2008a) among others.

3.6. **Fictional discourse.** The data we have seen so far may not speak directly to the issue
 430 of the empirical support for or against compositionality, but they point to the existence
 of richer meanings including world knowledge and co-speech gestures or, more generally,
 information derived from perceptual cues. Experimental research suggests that discourse
 is not only another source of content beyond strict, invariant lexical meanings, but can even
 add and subtract core features to elementary meanings themselves. One such extreme case
 435 of context-sensitivity can be found in fictional discourse. Nieuwland & van Berkum (2005)
 show that sentences which are otherwise sensible, like

(5) The peanut was salted.

appear anomalous if they are embedded in a context in which the inanimate subject ('the
 peanut') is attributed animate features. In a narrative in which the peanut danced, sang and
 440 met an almond it liked, 'salted' in (5) resulted in a *larger* N400 compared to 'in love' in (6)

(6) The peanut was in love.

This is taken to show that discourse can override even such deep-rooted semantic features
 as animacy. These findings can also be read as a challenge to Hintikka's principles, for they
 seem to show that meaning is context-dependent, and semantic composition can proceed
 445 from the outside-in, that is, from the discourse to lexical meaning. This may seem obvious
 to any observer of linguistic behavior, but the N400 gives that a twist: top-down effects can
 occur 'prior to the production of output' (Fodor 1983), which seems to threaten modularity.

Processing sentences such as (6) in a fictional context might therefore involve some form
 of top-down composition which, if it cannot resort to the mental lexicon as a repository of
 450 contextual values of expressions (recall Hodges' argument), then it must adopt some form
 of enriched composition – or give up a share of informational encapsulation, which is the
 same. However, there seems to be *another* way out for compositionality. One may ask, not
 what changes in the meaning of 'peanut' in a context in which it is depicted as animate, but
 what is preserved of the original (invariant, as we called it) meaning of the word. Animacy
 455 aside, there is no evidence that any of the other semantic features is maintained. Therefore,
 the word form 'peanut' in the fictional context considered here may just be used as a label

for an animate subject or, more precisely, a proper name with a reference but no (or perhaps very little) sense. This could easily be handled in a compositional manner. Processing the adjective ‘salted’, given the plausible combination with ‘peanut’, might recover its original
 460 sense, and this would explain the larger N400. This does not detract from the interest of the data, nor from the interpretation chosen by Nieuwland & van Berkum. It does exemplify, however, the kind of problems one encounters when trying to put compositionality to test, and in particular the exceptional elbow room compositionality leaves to its application to even the most controversial cases – it is precisely this resilience which has been taken by
 465 many as empirical vacuity.

3.7. **Semantic attraction.** A strict reading of compositionality implies that only two sorts of constraints can interact to produce complex meanings: syntax and the lexicon. A further assumption is that the syntax is an analysis of the sentence *as is given*, for instance a formal decomposition into constituents.¹⁰ This is all there is to the input for semantic composition.
 470 Combining lexical meanings (representations or types, depending on the ontology) based on the syntactic structure will in turn produce an input for the interpretation. In this sense, semantics is often said to be dependent on syntax. Kim & Osterhout (2005) designed an ERP study to test the extent to which syntax is actually in control of the composition process. They presented participants with sentences such as

- 475 (7) a. The hearty meal was devouring the kids.
 b. The hearty meal was devoured by the kids.

and found that ‘devouring’ in (7-a) elicited a larger P600 compared to ‘devoured’ in (7-b). If the syntax were taken as is given in the sentence – in well-formed sentences, that is – and if it were only proposing input to the semantics, an N400 to ‘devouring’ should be expected:
 480 indeed (7-a) is syntactically well-formed, whereas a semantic constraint (animacy) appears to be violated. The P600 indicates that (7-a) is perceived as a syntactic violation, originating from the impossibility at ‘devouring’ of building a passive construction. At the verb’s stem ‘devour-’, the passive is the only continuation compatible with a plausible meaning for the

¹⁰The extent to which a syntactic analysis is allowed to deviate from the surface form of an expression is a matter of considerable debate. For a discussion, see Culicover & Jackendoff (2005) and Culicover & Jackendoff (2006).

sentence, as is testified by (7-b). The data therefore show not only that semantic attraction to
 485 a more plausible interpretation is an important source of constraints in sentence processing
 – which could also be concluded if ‘devouring’ induced a larger N400 – but also that such
 constraints can override syntactic cues as these are given in the input – which is what the
 P600, as a syntax-related effect, shows. Compositionality can be salvaged only by assuming
 that semantic attractors, such as ‘kids devour hearty meals’, are configurations of the lexical
 490 network and not, as would seem more intuitively appealing, the result of inference. But this
 move is once again paradoxical: compositionality was introduced to explain productivity,
 and therefore to *ease* the burden on storage; now it seems we need a growing inventory of
 stored semantic facts to maintain compositionality.

3.8. **Coercion.** A phenomenon that has often been taken as a challenge to compositionality
 495 is complement coercion. Consider the following sentences

- (8) a. The journalist began the article after his coffee break.
 b. The journalist wrote the article after his coffee break.

The intuitive difference between the two is that, while ‘wrote the article’ asserts the relevant
 activity (writing), ‘began the article’ does not. So if a full event sense is to be recovered from
 500 (8-a), the activity must be inferred based on other semantic cues present in the sentence and
 stored knowledge. One candidate analysis (Pustejovsky 1995) focuses on the interpretation
 of the NP. For instance, ‘the article’ is an entity-denoting expression, which combined with
 verbs such as ‘begin’ denotes an event. Coercion is thus an instance of type-shifting, where
 types lay out a basic ontology of entities, events etc. An alternative analysis assumes richer
 505 event structures (van Lambalgen & Hamm 2004). Each VP is semantically represented by
 a quadruple $\langle f_1, f_2, e, f_3 \rangle$, where f_1 represents a force being exerted, f_2 the object or state
 driven by the force, e the goal toward which the exertion of the force is directed, and f_3 the
 state of having achieved that goal. Some slots in the quadruple may be empty, depending
 on the *Aktionsart* of the VP. Accomplishments such as ‘write an article’ feature a full event
 510 structure, while achievements such as ‘begin an article’ include only a punctual event e (the
 beginning of a yet unspecified activity relating to the article) and a consequent state (having
 begun the activity). Coercion is seen as a transition to a richer event structure, in which the

activity f_1 is also represented. Both analyses rely on some form of enriched composition, as in both cases an operation of meaning assembly that is not syntactic in nature (type-shifting
 515 or enrichment of the event structure) is postulated.

An interpretation of (8-a) in which the activity remains unspecified is conceivable, and it therefore falls to experimental research to provide evidence for or against the existence of complement coercion. A series of studies have shown that coercing sentences such as (8-a) result in increased processing costs compared to controls such as (8-b) (McElree et al. 2001;
 520 Traxler et al. 2002; Traxler et al. 2005; McElree et al. 2006; Pyllkkänen et al. 2007). This can be taken as evidence against strict composition.¹¹ However, a compositional analysis may still be possible if the operation which is responsible for generating the enriched meaning (say, type-shifting) is incorporated in the syntax. This choice can be criticised on different grounds. On empirical grounds, it predicts that (8-a) elicits a P600 (which, as we have seen,
 525 correlates with syntactic complexity), whereas the available neural data reveal a different effect with different neuronal sources (Pyllkkänen et al. 2007). On theoretical grounds, it leads to syntactic representations which resemble less and less a formal decomposition into constituents (Culicover & Jackendoff 2005): the simplicity and theoretical elegance which is gained by reintroducing compositionality is lost at the level of syntactic structure.

530

4. CONCLUSIONS

In this chapter we have tried to show that compositionality, properly operationalized, can be tested against processing data. We have also seen that behavioral and neurphysiological data undermine compositionality (simple composition), unless the balance between storage and computation is upset in favor of storage. It now seems appropriate to ask whether there
 535 is any interesting sense in which compositionality can be said to hold.

Compositionality (simple composition) remains effective as an explanation of cases in which processing complexity increases due to syntactic factors only. However, it falls short of accounting for situations in which complexity arises from interactions with the sentence or discourse context, perceptual cues and stored knowledge. The notion of compositionality
 540 as a methodological principle is appealing, but imputing the complexity to one component

¹¹Evidence for enriched composition in cases of aspectual coercion has also been found. See Piñango et al. (1999) and Piñango et al. (2006) among others.

of the grammar or other, instead of enriching the notion of composition, is not always an innocuous move, leading to fully equivalent theories. One may be tempted to believe that equivalent theories in this sense are *in principle* indistinguishable in the face of empirical data. However, neuroscience grants us (restricted) selective access to linguistic processes and representations in the brain, as exemplified by the difference between N400 and P600. 545 Therefore, there is at least a chance that what appear to be neutral methodological choices are in fact controvertible given empirical data. Compositionality sets also an upper bound on the degree of informational encapsulation posited by any modular or component-based theory of language. Parting from the upper bound, as the data might eventually force one to 550 do, implies weakening one's notion of composition, but also allowing more complex traffic either within a module or between modules. So compositionality is also crucial for issues of architecture and connectivity within the language system. Perhaps the most important of these issues is the balance between storage and computation: compositionality can often be rescued by taxing the storage component of the architecture, whereas it must be abandoned 555 if more demands on computation are placed.

Acknowledgments. We are grateful to the Netherlands Organization for Scientific Research for support under grant NWOCOG/04-19.

REFERENCES

- Baggio, G. & M. van Lambalgen (2007). The processing consequences of the imperfective paradox. *Journal of Semantics* 24, 307–330.
- Baggio, G., M. van Lambalgen, & P. Hagoort (2008a). Computing and recomputing discourse models: An ERP study. To appear.
- Baggio, G., M. van Lambalgen, & P. Hagoort (2008b). Language, linguistics and cognition. In M. Stokhof & J. Groenendijk (Eds.), *Handbook of Philosophy of Linguistics*. Amsterdam-New York: Elsevier. In preparation.
- Barton, S. & A. Sanford (1993). A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Memory and Cognition* 21, 477–487.
- Brown, C. & P. Hagoort (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience* 5, 34–44.

- Christianson, K., A. Hollingworth, J. Halliwell, & F. Ferreira (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42, 368–407.
- Croft, W. (1998). The structure of events and the structure of language. In M. Tomasello (Ed.), *The new psychology of language: cognitive and functional approaches to language structure*, pp. 67–92. Mahwah, NJ: Lawrence Erlbaum Associates.
- Culicover, P. & R. Jackendoff (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Culicover, P. & R. Jackendoff (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences* 10(9), 413–418.
- Dever, J. (1999). Compositionality as methodology. *Linguistics and Philosophy* 22, 311–326.
- Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In *Evoked Brain Potentials and Behavior*, pp. 13–88. Plenum Press.
- Erickson, T. & M. Matteson (1981). From words to meanings: A semantic illusion. *Journal of Verbal Learning and Verbal Behaviour* 20, 540–552.
- Ferreira, F., V. Ferraro, & K. Bailey (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11, 11–15.
- Ferreira, F. & N. Patson (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass* 1, 71–83.
- Ferretti, T., M. Kutas, & K. McRae (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33, 182–196.
- Fodor, J. (1983). *The Modularity of Mind*. MIT Press.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance*, Volume XII, pp. 559–586. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frege, G. (1884). *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Koebner.
- Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78–84.
- Groenendijk, J. & M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.

- Hagoort, P. (2003). How the brain solves the binding problem for language: A neuro-computational model of syntactic processing. *Neuroimage* 20, S18–S29.
- Hagoort, P. & C. Brown (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing*, pp. 45–81. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hagoort, P., C. Brown, & J. Groothusen (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8, 439–483.
- Hagoort, P., C. Brown, & L. Osterhout (2001). The neurocognition of syntactic processing. In *The Cognitive Neuroscience of Language*, pp. 273–307. Oxford University Press.
- Hagoort, P., L. Hald, M. Bastiaansen, & K. Petersson (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441.
- Hintikka, J. (1983). *The Game of Language. Studies in Game-Theoretical Semantics and its Applications*. Synthese Language Library. Reidel.
- Hodges, W. (2006). From sentence meanings to full semantics. Unpublished manuscript.
- Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press.
- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Janssen, T. (1986). *Foundations and Applications of Montague Grammar, Part 1*. Amsterdam: Center for Mathematics and Computer Science.
- Janssen, T. (1997). Compositionality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of Logic and Language*. Elsevier Science.
- Katz, J. & J. Fodor (1963). The structure of a semantic theory. *Language* 39(2), 170–210.
- Kazmi, A. & F. Pelletier (1998). Is compositionality formally vacuous? *Linguistics and Philosophy* 21, 629–633.
- Keenan, E. (1979). On surface form and logical form. *Studies in the Linguistic Sciences* 8, 163–203.
- Kim, A. & L. Osterhout (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language* 52, 205–225.
- Kutas, M. & S. Hillyard (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207, 203–205.

- Kutas, M. & S. Hillyard (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- McElree, B., L. Pyllkanen, M. Pickering, & M. Traxler (2006). A time course analysis of enriched composition. *Psychonomic Bulletin and Review* 13(1), 53–59.
- McElree, B., M. Traxler, M. Pickering, R. Seely, & R. Jackendoff (2001). Reading time evidence for enriched composition. *Cognition* 78(0010-0277 (Print)), B17–25.
- Mecklinger, A., H. Schriefers, K. Steinhauer, & A. Friederici (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition* 23, 477–494.
- Miller, G. & P. Johnson-Laird (1976). *Language and Perception*. Harvard University Press.
- Müller, H., J. King, & M. Kutas (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research* 5, 193–203.
- Münte, T., K. Schiltz, & M. Kutas (1998). When temporal terms belie conceptual order. *Nature* 395, 71–3.
- Nieuwland, M. & J. van Berkum (2005). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18, 1098–1111.
- Osterhout, L. & P. Holcomb (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31, 785–806.
- Özyürek, A., R. Willems, S. Kita, & P. Hagoort (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19, 605–616.
- Partee, B., A. ter Meulen, & R. Wall (1990). *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.
- Piñango, M., A. Winnick, R. Ullah, & E. Zurif (2006). The time course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research* 35, 233–244.

- Piñango, M., E. Zurif, & R. Jackendoff (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research* 28, 395–414.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pylkkänen, L., R. Llinas, & B. McElree (2007). An MEG study of silent meaning. Submitted.
- Sanford, A. (2002). Context, attention and depth of processing during interpretation. *Mind and Language* 17, 188–206.
- Sanford, A. & P. Sturt (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences* 6, 382–386.
- Traxler, M., B. McElree, R. Williams, & M. Pickering (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language* 53, 1–25.
- Traxler, M., M. Pickering, & B. McElree (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language* 47, 530–547.
- Van Berkum, J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond*, pp. 229–270. New York: Psychology Press.
- van Berkum, J., P. Hagoort, & C. Brown (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience* 11, 657–671.
- van Berkum, J., P. Zwitserlood, P. Hagoort, & C. Brown (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research* 17, 701–718.
- van Lambalgen, M. & F. Hamm (2004). *The Proper Treatment of Events*. Oxford: Blackwell.
- van Oostendorp, H. & S. de Mul (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica* 74, 35–46.
- Westerståhl, D. (1998). On the mathematical proofs of the vacuity of compositionality. *Linguistics and Philosophy* 21, 635–643.

- Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy* 17, 329–342.

Computing and recomputing discourse models: An ERP study

Giosuè Baggio^{a,*}, Michiel van Lambalgen^b, Peter Hagoort^{c,d}

^a *F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands*

^b *Institute for Logic, Language and Computation University of Amsterdam, The Netherlands*

^c *F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands*

^d *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Received 11 December 2007; revision received 11 February 2008

Available online 28 April 2008

Abstract

While syntactic reanalysis has been extensively investigated in psycholinguistics, comparatively little is known about reanalysis in the semantic domain. We used event-related brain potentials (ERPs) to keep track of semantic processes involved in understanding short narratives such as ‘The girl was writing a letter when her friend spilled coffee on the paper’. We hypothesize that these sentences are interpreted in two steps: (1) when the progressive clause is processed, a discourse model is computed in which the goal state (a complete letter) is predicted to hold; (2) when the subordinate clause is processed, the initial representation is recomputed to the effect that, in the final discourse structure, the goal state is not satisfied. Critical sentences evoked larger sustained anterior negativities (SANs) compared to controls, starting around 400 ms following the onset of the sentence-final word, and lasting for about 400 ms. The amplitude of the SAN was correlated with the frequency with which participants, in an offline probe-selection task, responded that the goal state was not attained. Our results raise the possibility that the brain supports some form of non-monotonic recomputation to integrate information which invalidates previously held assumptions.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Semantics; Discourse; Non-monotonicity; ERPs; Sustained anterior negativity

Introduction

In the past three decades experimental research using event-related potentials (ERPs) has provided numerous insights into word, sentence and discourse comprehension. However, as has been noted, “a cognitive neuroscience approach to language has not as yet merged with linguistic and psycholinguistic research programmes”

(Brown & Hagoort, 1999). One linguistic research program that may contribute to understanding the basis of meaning in the human brain is semantic theory. Logicians and formal semanticists since the ‘dynamic turn’ (Peregrin, 2003) have shifted their attention from describing semantic competence to modeling cognitive update and information exchange. A case in point is a recent proposal by van Lambalgen and Hamm (2004) which regards comprehension as an incremental, yet non-monotonic process whereby temporary structures are set up in working memory and may be later revised on the basis of further discourse information. Although

* Corresponding author. Fax: +31 024 3610989.

E-mail address: g.baggio@fcdonders.ru.nl (G. Baggio).

evidence for semantic reanalysis exists (Carreiras, Garnham, Oakhill, & Cain, 1996; Sturt, 2007), the issue has arguably received less attention than it deserves. The purpose of the present study is to contribute filling this gap. We used ERPs to test a processing hypothesis proposed by Baggio and van Lambalgen (2007) as an application of the formal, non-monotonic semantics of tense and aspect of van Lambalgen and Hamm (2004).

ERP research on semantic processing

Event-related potentials have proved useful to address a number of issues concerning the relative complexity and time course of semantic processes. Kutas and Hillyard (1980) conducted the first ERP experiment in which linguistic factors were successfully manipulated, in this case the semantic plausibility of a word given the preceding sentence context:

- (1) a. The officer shot the man with a gun
b. The officer shot the man with a moon

Compared to ‘gun’, the anomalous noun ‘moon’ resulted in a larger negative shift starting around 250 ms after word onset, peaking at 400 ms, and lasting for approximately another 150 ms. This ERP component, called N400 after its polarity and peak latency, is known not to be affected by other unexpected events, such as variations in the physical properties of the stimuli. Larger N400s are also triggered by semantically plausible words which are nevertheless judged as less preferred in a given sentence context (Hagoort & Brown, 1994; Kutas & Hillyard, 1984), for example ‘pocket’ in (2b):

- (2) a. Jenny put the sweet in her mouth after the lesson
b. Jenny put the sweet in her pocket after the lesson

The amplitude of the N400 is also modulated by lexical items which provide information conflicting with the discourse context (van Berkum, Hagoort, & Brown, 1999; van Berkum, Zwitserlood, Hagoort, & Brown, 2003) or world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004). In sum, although every content word evokes an N400, the amplitude of the negative shift appears to be affected by the degree of semantic fit of a lexical item with the preceding context and the knowledge base relevant for its integration.

Semantics-related negativities different from the N400 have also been found. Van Berkum and colleagues recorded ERPs while subjects read (van Berkum, Brown, & Hagoort, 1999) and listened to (van Berkum, Brown, Hagoort, & Zwitserlood, 2003) discourses in which a particular NP in a target sentence could denote either

a single referent introduced in the preceding discourse or two equally suitable referents. For instance, (3c), containing the NP ‘the girl’, could follow either the single-referent context (3a) or the double-referent context (3b):

- (3) a. David had told the boy and the girl to clean up their room before lunch time. But the boy had stayed in bed all morning, and the girl had been on the phone all the time
b. David had told the two girls to clean up their room before lunch time. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time
c. David told the girl that...

Referentially ambiguous NPs, such as ‘the girl’ in (3c) following (3b), elicited a sustained anterior negativity (SAN), emerging 300–400 ms after noun onset and lasting for several hundreds of milliseconds. The SAN differed from typical instances of the N400 in duration (‘sustained’) and scalp distribution (‘anterior’). Following earlier claims (Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Müller, King, & Kutas, 1997), the time-course and topographical profile of the observed ERPs are taken to suggest that “at least some of the processing consequences of referential ambiguity may involve an increased demand on memory resources” (van Berkum et al., 2003).

Another study (Münte, Schiltz, & Kutas, 1998) reported sustained anterior negativities. ERPs were recorded while subjects read narratives differing in the initial temporal connective:

- (4) a. After the scientist submitted the paper, the journal changed its policy
b. Before the scientist submitted the paper, the journal changed its policy

‘Before’ sentences elicited a larger sustained negativity, maximal over left anterior sites. At the left frontal electrode, ERP responses to ‘before’ and ‘after’ diverged approximately 300 ms after sentence onset. The effect lasted throughout the sentence and was larger during the second clause. The difference of anterior negativity between ‘before’ and ‘after’ items was positively correlated with participants’ working memory span. Münte et al. argue that the slow negative shift evoked by ‘before’ sentences reflects working memory operations involved in computing a model for (4b) in which the events are represented in their actual order of occurrence. That is, in contrast with (4a), (4b) requires *additional* memory resources as the two events are mentioned in reverse temporal order.

The connection between sustained anterior negativities and working memory is explicit in van Berkum et al. (1999), van Berkum et al. (2003) and Münte

et al. (1998). However, there is no full agreement on a functional account, based on linguistically-informed notions, of these findings. For instance, while van Berkum et al. suggest that the sustained anterior negativity reflects ‘referential processing’, Münte et al. seem to implicate that ‘additional discourse-level computations’ of the temporal and causal profiles of the events described by ‘before’ and ‘after’ sentences are responsible for the observed slow negative shifts. Matters appear to be further complicated by the finding that sustained anterior negativities are elicited by constructions in which complexity at the syntax-semantics interface is affected, as in long-distance *wh*-dependencies (Felsler, Clahsen, & Münte, 2003; Fiebach, Schlesewsky, & Friederici, 2002; King & Kutas, 1995; Müller et al., 1997; Phillips, Kazanina, & Abada, 2005).

Despite the differences between the conditions in which sustained anterior negativities have been observed, the proposed processes can be brought under a single umbrella term, which we shall refer to as ‘computing a discourse model’. Formal semantics, at least since Discourse Representation Theory (DRT) (Kamp, 1981), has assumed that interpreting definite and indefinite NPs, resolving anaphoric pronouns, determining the order of events, establishing *wh*-dependencies and other cross-clause and cross-sentence processes concur in the construction of a discourse model, that is, a cognitive representation making a given narrative true. More recent proposals, which build upon DRT and add some sophistication to it, regard discourse comprehension as a process in which lexical meanings, context and world knowledge interact to produce consistent discourse representations (Hamm, Kamp, & van Lambalgen, 2006; van Lambalgen & Hamm, 2004). Pragmatic constraints and causal/world knowledge are brought to the fore by these accounts. Furthermore, discourse models as envisaged by the theory (called ‘minimal models’, see “Minimal models, inference in the limit and recomputation” section) can be efficiently computed by artificial neural networks, which account for some capabilities and limitations of working memory (Stening & van Lambalgen, 2005). Therefore, in this framework it becomes possible to raise and address a number of issues concerning the complexity of computing discourse models in working memory. To see in some detail how this could be done, we must first introduce the linguistic phenomenon with which we shall be concerned: the ‘imperfective paradox’.

The imperfective paradox

Verb phrases (VPs) can be semantically classified as states (‘know’, ‘love’, etc.), activities (‘run’, ‘write’, etc.), accomplishments (‘write a letter’, ‘bake a cake’, etc.), achievements (‘finish’, ‘reach’, etc.) and points (‘flash’, ‘hop’, etc.) (Steedman, 1997). Accomplishments

involve the activity from which they are derived. For instance, ‘write a letter’ is constituted by the activity ‘write’ and the direct object ‘a letter’, which need not refer to an existing entity, but carries information about the goal toward which the writing activity is directed. Here we use the term ‘activity’ to denote both the aspectual class of VPs such as ‘write’ in the above classification and the atelic process involved in all accomplishments. We use ‘progressive’ and ‘imperfective’ interchangeably to allow the reader to see the connection between the semantics of the progressive and the imperfective paradox, although this is not entirely correct (Comrie, 1976).

Let us consider accomplishments first:

- (5) The girl was writing a letter when her friend spilled coffee on the tablecloth.

From (5) the reader would typically conclude that, barring unforeseen circumstances, the girl will attain the desired goal and would thus assent to the statement ‘The girl has written a letter’ (see Sections “Entailment questionnaire” and “Behavioral data” for evidence supporting this claim). Such an inference is based on the assumption that spilling coffee on the tablecloth is usually *neutral* with respect to the writing activity, that is, it is not a typical immediate cause leading to the termination of the activity. It is possible to imagine situations in which writing was temporarily interrupted or even terminated by the accident. However, as the data reported in Sections “Entailment questionnaire” and “Behavioral data” will demonstrate, failing to explicitly mention an obstacle in the discourse is sufficient to lead the reader to assume that there was no such obstacle to attaining the intended goal.

We hypothesize that the inference to a goal state is *defeasible* or *non-monotonic*, that is, it can be suppressed if the discourse describes an event which terminates the relevant activity:

- (6) The girl was writing a letter when her friend spilled coffee on the paper.

Assuming that writing was intended to occur on the same paper sheets on which coffee was spilled, the accident is sufficient to terminate the activity and it is therefore a *disabling* condition for obtaining a complete letter. Accordingly, on the basis of (6) the reader would assent to ‘The girl has written no letter’.

Suppression can obtain only with accomplishments, not with activities (Rothstein, 2004). In accomplishments, the object NP ‘a letter’ expresses the existence of a natural culmination point or ‘canonical goal’ toward which the writing activity is directed, namely a complete letter. Activities, for instance ‘writing letters’, do not involve any such canonical goal. The use of the

bare plural ‘letters’ indicates that the number of letters is (for the speaker and the hearer) unspecified and that, therefore, the activity has no natural culmination point. Accordingly, a narrative containing the activity VP ‘writing letters’ will be interpreted as entailing that ‘The girl has written one or more letters’ regardless of the consequences of the second event on the writing activity:

- (7) The girl was writing letters when her friend spilled coffee on the tablecloth.
- (8) he girl was writing letters when her friend spilled coffee on the paper.

There appears to be something paradoxical about (6) in its relation to (5), which is not found in the pair (7) and (8). Whereas it belongs to the meaning of the accomplishment ‘writing a letter’ that the writing activity is directed toward the goal state of a complete letter, the actual occurrence of that consequent state can be denied without contradiction. How can an essential component of the meaning be denied without destroying meaning itself? This is the so-called ‘imperfective paradox’. We now turn to its processing consequences.

Minimal models, inference in the limit and recomputation

Language processing amounts to incrementally computing a discourse representation given lexical, syntactic and contextual constraints (Hagoort, 2006). To make computation tractable, discourse models must be ‘minimal’, that is, in a precise mathematical sense (van Lambalgen & Hamm, 2004), the simplest possible structures making the narrative true. Minimal models behave like ‘closed worlds’, in which only those propositions which are asserted in discourse, or which can be inferred from it or from background knowledge, are represented as true in the model. For the remaining cases, a distinction must be drawn. Propositions which are mentioned in discourse, but are not asserted and do not follow from what is said or from background knowledge (e.g. the antecedent of a conditional), are represented as false in the minimal model. In logical terms, these propositions still belong to the finite language upon which the construction of the minimal model is based. But as long as nothing forces their truth, they will be taken as false. Propositions which are not part of the finite language—because they do not occur in the discourse context or in background knowledge—are not included in the minimal model, that is, they are not represented as being either true or false.

One important upshot of the theory is that the occurrence of a goal state can be inferred from a minimal model of a discourse containing an accomplishment in the past progressive. As soon as the sentence ‘The girl was writing a letter’ is processed,

the system constructs a minimal model in which the goal state (a complete letter) is attained at some time later than the interval referred to by the progressive. Two remarks concerning this crucial point are in order. First, interpretation is based on the ‘closed world assumption’: if no disabling condition is described in discourse (so far), it will be (temporarily) assumed that there is no obstacle interfering with the writing activity. Second, the conclusion that eventually a letter is accomplished is an instance of *predictive inference* or, more precisely, *inference in the limit*: given that writing is asserted to hold some time in the past, that it can be assumed there are no obstacles for the writing activity, that some form of inertia holds (writing continues if it is not hindered by external forces) and that a letter is a finite object, it can be expected that the process will converge—‘in the limit’—to a complete letter. This holds for both neutral (5) and disabled (6) accomplishments. Now, when the initial model is extended with a ‘when’ clause describing an event which terminates the writing activity (i.e. a disabling condition), the goal state inference will be suppressed. The subordinate clause ‘when her friend spilled coffee on the paper’ will lead to the retrieval of causal knowledge from semantic memory to the effect that the coffee accident terminated the writing activity. Spilling occurred *during* the writing process, from which follows that the accident took place *before* a complete letter was obtained. The writing event can be imagined as an open interval, where the goal state (a complete letter) is no longer part of the structure. We shall use the term ‘recomputation’ to refer to the suppression of the goal state inference when the subordinate clause in (6) is processed. Because (5) describes a neutral scenario, the goal state derived while processing the progressivized VP is maintained in the final model. In conclusion, whereas (5) involves an *extension* of the initial discourse model, (6) might induce a *recomputation*. Since (7) and (8) do not involve a canonical goal, they will require an extension only.

Predictions for ERPs

The only difference between neutral and disabled *activities* (e.g. ‘writing letters’) is the noun in the subordinate clause, ‘tablecloth’ or ‘paper’. In both cases the initial model is simply extended, thus we expect to observe only local ERP differences related to the integration of the differing nouns. As ‘tablecloth’ is less semantically related to the other lexical items occurring in the sentence compared to ‘paper’, we expect a larger N400 for the former compared to the latter word.

Processing a ‘when’ clause following an *accomplishment* (e.g. ‘writing a letter’) involves integrating the differing nouns *and*, in the disabling case, recomputing

the initial discourse representation. Also in this case, the neutral noun ‘tablecloth’ is predicted to evoke a larger N400 compared to the disabling ‘paper’, reflecting a lower degree of semantic relatedness with the preceding context. In our ERP study Dutch materials were used, where the verb in subordinate clauses occupies the sentence-final position (see “Materials” section). The temporal and causal information provided by verbs in ‘when’ clauses is necessary to initiate the recomputation process (Baggio & van Lambalgen, 2007). Thus, the ERP effects of what we have analyzed as recomputation are expected to surface at the sentence-final verb ‘spilled’ (‘morste’ in our Dutch stimuli, see “Materials” section and Table 1).

One additional prediction is that the amplitude of the ERP effect evoked by disabled accomplishments is correlated with the relative frequency with which readers infer that the goal state was not attained. Recomputation is expected to consistently evoke a time-locked shift in the EEG in each trial in which a negative judgment concerning the attainment of the goal is made. Therefore, the higher the frequency

of such inferences—the larger the number of trials in which recomputation occurred—the larger the amplitude of the ERP component. The method and results of an ERP study in which these predictions were tested are described below.

Methods

Materials

The set of Dutch materials used in the experiment included 160 test and 160 filler items (see Supplementary Material). Each test item included two context sentences providing a neutral setting for the events, four target sentences (A)–(D) and two probe pairs (E) and (F) (Table 1). Target sentences were constructed by manipulating the aspectual class of the progressive VP (activity or accomplishment) and the effects of the event introduced by the ‘when’ clause (neutral or disabling) on the event described in the main progressive clause. All progressive VPs were instances of the Dutch periphrastic

Table 1
Examples of stimulus sentences

<i>Context sentences</i>	
	De deur van de woonkamer was gesloten. Binnen speelde de radio klassieke muziek. <i>The door of the living-room was closed. Inside played the radio classical music.</i> ^a ‘The door of the living room was closed. Inside the radio played classical music’. ^b
<i>Target sentences</i>	
(A)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed morste. <i>The girl was letters on the to-write when her friend coffee on the tablecloth spilled.</i> ^a ‘The girl was writing letters when her friend spilled coffee on the tablecloth’. ^b
(B)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier morste. <i>The girl was letters on the to-write when her friend coffee on the paper spilled.</i> ^a ‘The girl was writing letters when her friend spilled coffee on the paper’. ^b
(C)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste. <i>The girl was a letter on the to-write when her friend coffee on the tablecloth spilled.</i> ^a ‘The girl was writing a letter when her friend spilled coffee on the tablecloth’. ^b
(D)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste. <i>The girl was a letter on the to-write when her friend coffee on the paper spilled.</i> ^a ‘The girl was writing a letter when her friend spilled coffee on the paper’. ^b
<i>Probe sentences</i>	
(E)	Het meisje heeft een of meer brieven geschreven. <i>The girl has one or more letters written.</i> ^a ‘The girl has written one or more letters’. ^b Het meisje heeft geen brief geschreven. <i>The girl has no letter written.</i> ^a ‘The girl has written no letter’. ^b
(F)	Het meisje heeft een brief geschreven. <i>The girl has a letter written.</i> ^a ‘The girl has written a letter’. ^b Het meisje heeft geen brief geschreven. <i>The girl has no letter written.</i> ^a ‘The girl has written no letter’. ^b

^a Literal translation.

^b Paraphrase.

‘was/waren NP aan het V_{inf} ’ construction. This solution is to be preferred to the use of the Dutch simple past which, in some cases, is aspectually ambiguous between perfective and imperfective readings. Accomplishments differed from activities in the object NP only: an indefinite (‘een brief’/‘a letter’) was used for accomplishments, a bare plural (‘brieven’/‘letters’) for activities. Disabling and neutral subordinate clauses differed only in the prepositional or object NP, for instance ‘papier’ and ‘tafelkleed’. Neutral and disabling events were distinguished based on the experimenters’ judgment (but see “Entailment questionnaire” section for some data supporting these choices). Probe pairs (E) were used with activities and (F) with accomplishments (Table 1).

Fillers were 160 sentences of varying length, structure and content. Analogously to test items, fillers were preceded by two neutral context sentences and followed by a probe pair. Target sentences described an event consistently, as in (9), or inconsistently, as in (10), with factual knowledge (see (Hagoort et al. (2004) for an experiment based on these stimuli):

- (9) Dutch trains are white and very crowded.
 (10) Dutch trains are yellow and very crowded.

Probes were of the type ‘Trains in the Netherlands are white.’/‘Trains in the Netherlands are yellow.’ These fillers were chosen to add variety to the materials while preserving the task used for test items.

Four test versions were constructed, comprising randomized lists of test and filler items. The task was identical for critical and filler sentences. Participants had to select the correct probe based on the information provided by the context and target items. Mean length, raw and lemma frequency of the differing nouns in the NP of subordinate clauses were matched using the CELEX Dutch corpus (Baayen, Piepenbrock, & Gulikers, 1996). Mean length was 7.9 letters ($SD = 2.46$) for neutral and 7.75 ($SD = 2.79$) for disabled cases, and was kept below 12 letters in any case. Raw frequency per million words was 1113 ($SD = 2462$) for neutral and 1096 ($SD = 2792$) for disabled cases. Lemma frequency per million words was 1730 ($SD = 3559$) for neutral and 1666 ($SD = 3585$) for disabled cases. The length of sentence-final verbs was identical across conditions and was kept below 12 letters in any case. Cloze probabilities of sentence-final verbs were normed in a dedicated pre-test discussed below.

Pre-tests

Cloze probability test

In order to determine the cloze probabilities of sentence-final verbs, context sentences followed by a target sentence with the final word blanked were presented to a group of 32 native speakers of Dutch (mean age 22.5, 27

female). Participants were requested to fill in the blank with the first word that came to their mind. Four versions (40 items per condition), randomized and balanced across conditions, were constructed. Mean cloze probabilities were not different between the conditions (all comparisons using t -tests, $P > .05$) in each test version as well as in the entire set.

Entailment questionnaire

A paper-and-pencil judgment task was also administered. Thirty six Dutch native speakers (mean age 22.5, 28 female) were presented with the context followed by a target sentence and a probe pair. The task was to select the appropriate probe. Negative probes were more frequently chosen for disabled accomplishments than for the other conditions. Neutral activities (S1) (see Table 1) showed the lowest mean of negative responses ($M = 2.72$, $SD = 3.22$), followed by disabled activities (B) ($M = 8.06$, $SD = 7.05$), neutral accomplishments (C) ($M = 1.03$, $SD = 9.23$) and disabled accomplishments (D) ($M = 25.14$, $SD = 8.02$) (see Baggio & van Lambalgen, 2007 for details).

Participants

Thirty-one students participated in the ERP experiment. Of these, seven were left out of the final analysis due to a high number (>20%) of trials contaminated by artifacts. This left us with twenty four participants (mean age 22.5, 17 female), with no history of neurological, psychiatric or cognitive disorders. Subjects were selected from the database of the F.C. Donders Centre for Cognitive Neuroimaging at the Radboud University Nijmegen. Participants received €8 per hour or course credits. None of the subjects who took part to the pre-tests participated in the ERP experiment.

Procedure

After applying the electrodes (see “Recording” section), participants were conducted into the experimental room and were asked to sit in front of a video monitor. The stimuli were presented as follows: the two context sentences were displayed together on a single screen (white on black background) for a variable duration (6, 7 or 8 s), depending on the length of the sentences themselves; next the target sentence, one of (A)–(D), was presented on the screen word-by-word (600 ms SOA, 300 ms word duration; white on black background); the target sentence was preceded and followed by a fixation cross, presented for 1500 ms; finally, the probe pair, one of (E) and (F), was shown on the screen (red on black background) and remained visible until the participant gave a button-press; the probes were followed by a fixation cross which lasted for 1500 ms. The same presentation parameters were used for fillers. Participants were

instructed to read each sentence carefully, to blink only when the fixation cross was shown and to select the correct probe by pressing one of two buttons (left or right on the button box) as quickly and accurately as possible. The position on the screen (top or bottom) of the positive and negative probe corresponded to the left and right button respectively, and was counterbalanced across test versions. In this way, participants could not prepare their motor response before the probe pair was presented on the screen. The experiment took about 2 h and was divided into 24 blocks of 10 trials each.

Recording

EEG and EOG signals were recorded using Ag/AgCl electrodes. The EOG was measured from four electrodes: one at the outer canthus of each eye, one below and one above the left eye (FE). The EEG was measured from 28 electrodes, arranged according to American Electrophysiological Society conventions: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, and O2. Two additional electrodes were placed on the left and right mastoids, the former serving as the reference during the measurement. All EEG and EOG electrodes were referenced offline to a linked mastoid. EEG electrodes were attached to an elastic cap, whereas EOG and reference electrodes were applied using two-sided adhesive decals external to the cap. Electrode impedance was kept below 5 k Ω throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with a 500-Hz sampling rate, a low pass filter set at 125 Hz and a 10-s time constant.

Data analysis

Data analysis was conducted using FieldTrip,¹ a MATLAB package for processing EEG signals. The following transforms were applied to each subject's dataset. Segments corresponding to the noun and the sentence-final verb were extracted from the EEG with an interval of 200 ms preceding and 800 ms following stimulus onset. Baseline correction used the 200-ms interval preceding the onset of nouns, and the 100-ms interval following the onset of sentence-final verbs. The latter choice was effected so as to prevent ERP differences in the 400- to 600-ms interval following the onset of the nouns from biasing the baseline correction for the ERPs evoked by sentence-final verbs in the same time interval. The use of such a baseline seems acceptable on grounds that the expected recomputation effect at the verb would not affect such largely exogenous components as the N1.

Artifact rejection was based on two FieldTrip functions: the first detects and rejects all trials that contain activity exceeding a threshold of $\pm 100 \mu\text{V}$; the second identifies and discards trials contaminated with eye movements or blinks by means of thresholding the z -transformed value of the raw data in the EOG channels, preprocessed using a band-pass filter of 1–15 Hz. A 30-Hz low-pass digital filter was applied to the segmented, artifact-free data. ERPs were obtained for each subject by averaging over trials in each experimental condition. A 5-Hz low-pass filter was used to produce the waveforms shown in Figs. 2–5. Topographical plots and statistical analyses were however based on the 30-Hz low-pass filtered data.

For the analysis of behavioral responses we employed two repeated-measures ANOVA models with Subject as the random effect, Aspectual Class (Activity/Accomplishment) and Subordinate Clause Type (Neutral/Disabling) as fixed effects, and the mean value of either negative judgments (negative probes selected in the response task) or decision times in each condition as the dependent variables.

Statistical analyses of ERP data used a non-parametric randomization procedure (Maris, 2004; Maris & Oostenveld, 2007) which took as input mean amplitude (μV) values in each condition in time bins of 100 ms, starting from the onset of the relevant word and ending 800 ms after, and produced as output a cluster of electrodes (min. 1 and max. 28) in which the difference between the conditions was significant in each time bin, the sum of t -statistics in that cluster and Monte Carlo estimates of P -values.

For the correlation analysis (see Section Predictions for ERPs), we calculated the difference between the ERPs evoked by sentence-final verbs in subordinate clauses—disabled (D) minus neutral (C)—following accomplishments at anterior sites (Fp1, Fp2, F7-, F3-, Fz-, F4-, F8-averaged) in the 500- to 700-ms interval after the onset of the sentence-final verb (see “Event-related brain potentials” section for motivation). Pearson's product-moment correlation was computed to determine whether the amplitude difference in ERPs varied with the number of negative responses, quantified again as the difference of negative judgments between disabled (D) and neutral (C) accomplishments. The correlation analysis was done on a per-subject basis (i.e. each pair of data points in the correlation corresponded to a single subject's data).

Results

Behavioral data

Neutral activities (S1) showed the lowest mean of negative responses ($M = 4.08$, $SD = 2.87$), followed by disabled activities (B) ($M = 5.83$, $SD = 4.51$), neutral

¹ For more information, see <http://www2.ru.nl/fcdonders/fieldtrip/>.

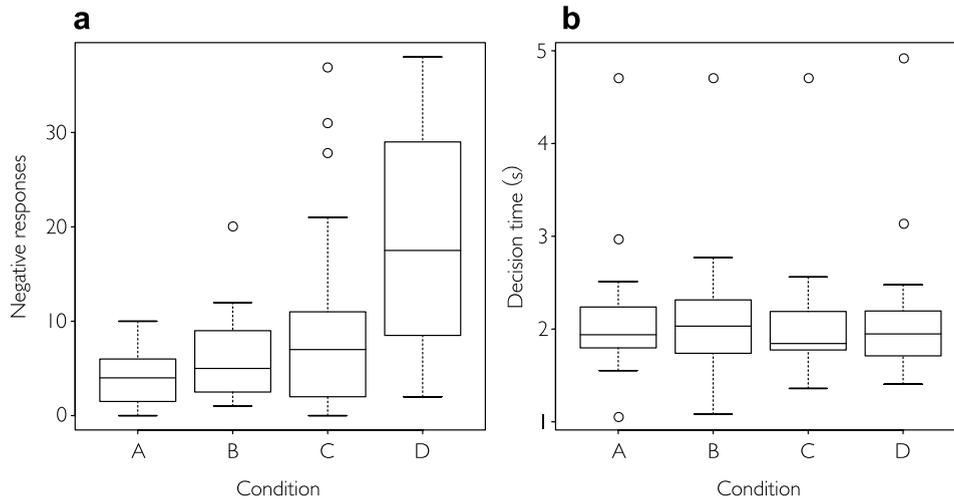


Fig. 1. Behavioral data. (a) Boxplot of the categorical responses. (b) Boxplot of the decision times. Conditions are represented on the abscissa (see Table 1 for the corresponding labels). Negative responses and decision times are plotted on the ordinate. The solid line within the boxes indicates the median, box height is equal to the interquartile range, whiskers represent adjacent values, and empty circles denote outliers. The maximum of potential negative responses is 40.

accomplishments (C) ($M = 9.58$, $SD = 9.96$) and disabled accomplishments (D) ($M = 18.13$, $SD = 11.16$). The distribution of the data in the different conditions appears rather similar, as indicated by box height and whisker length in Fig. 1a. However, disabled accomplishments have a more spread-out distribution, suggesting that inference patterns were less uniform across participants. ANOVAs revealed significant main effects of Aspectual Class and Subordinate Clause Type, and a significant interaction between the two factors (Table 2, Fig. 1a). The observed pattern of responses supports the linguistic views outlined above and replicates our previous findings (see “Entailment questionnaire” section). There is no difference in decision times (Table 2, Fig. 1b): (S1), $M = 2111$ ms, $SD = 677$ ms; (B), $M = 2100$ ms, $SD = 680$ ms; (C), $M = 2070$ ms, $SD = 646$ ms; (D), $M = 2086$ ms, $SD = 710$ ms.

Event-related brain potentials

Nouns

Fig. 2 displays the ERP topographies and waveforms elicited by *activities*. An N1 component peaking at

Table 2
Summary of ANOVA statistics for behavioral data

	Categorical responses	Decision times
Aspectual class	$F(1, 23) = 21.65$ $P < .001$	$F(1, 23) < 1$
Subordinate clause type	$F(1, 23) = 23.60$ $P < .001$	$F(1, 23) < 1$
Aspectual class \times Subordinate clause type	$F(1, 23) = 17.20$ $P < .001$	$F(1, 23) < 1$

approximately 100 ms is followed by a P2 component with a trough at about 200 ms. The amplitude of the N1 and F is not different between disabled and neutral clauses: no significant clusters were found between 0 and 300 ms (all contrasts, $P > 0.1$). The N1–P2 complex is followed by an N400. The amplitude of the N400 is larger in neutral (‘tafelkleed’) than in disabling (‘papier’) clauses (Fig. 2b): significant clusters with a central distribution were found between 300 and 500 ms (Table 3, Fig. 2a). No difference between neutral and disabling clauses was found after 500 ms.

Fig. 3 displays the ERP topographies and waveforms elicited by *accomplishments*. Also in this case, an N1–P2 complex can be observed. There is no difference between neutral and disabling clauses, as no significant clusters between 0 and 300 ms were found (all contrasts, $P > 0.1$). The amplitude of the N400 is again larger in neutral (‘tafelkleed’) than in disabling (‘papier’) clauses (Fig. 3b). The effect lasts longer than the N400 observed in activities: significant clusters with a central distribution were found between 300 and 600 ms (Table 3 and Fig. 3a). No difference between conditions was found after 500 ms.

There is no overall difference between the two aspectual classes. Cluster-based t -tests comparing the N400 effects in activities and accomplishments (corresponding to testing the main effect of Aspectual Class in a parametric model) produced no significant clusters between 300 and 600 ms from noun onset (all contrasts, $P > 0.1$). No difference was found in any of the remaining time bins.

Sentence-final verbs

Fig. 4 displays the ERP topographies and waveforms elicited by *activities*. Contrary to what we had observed

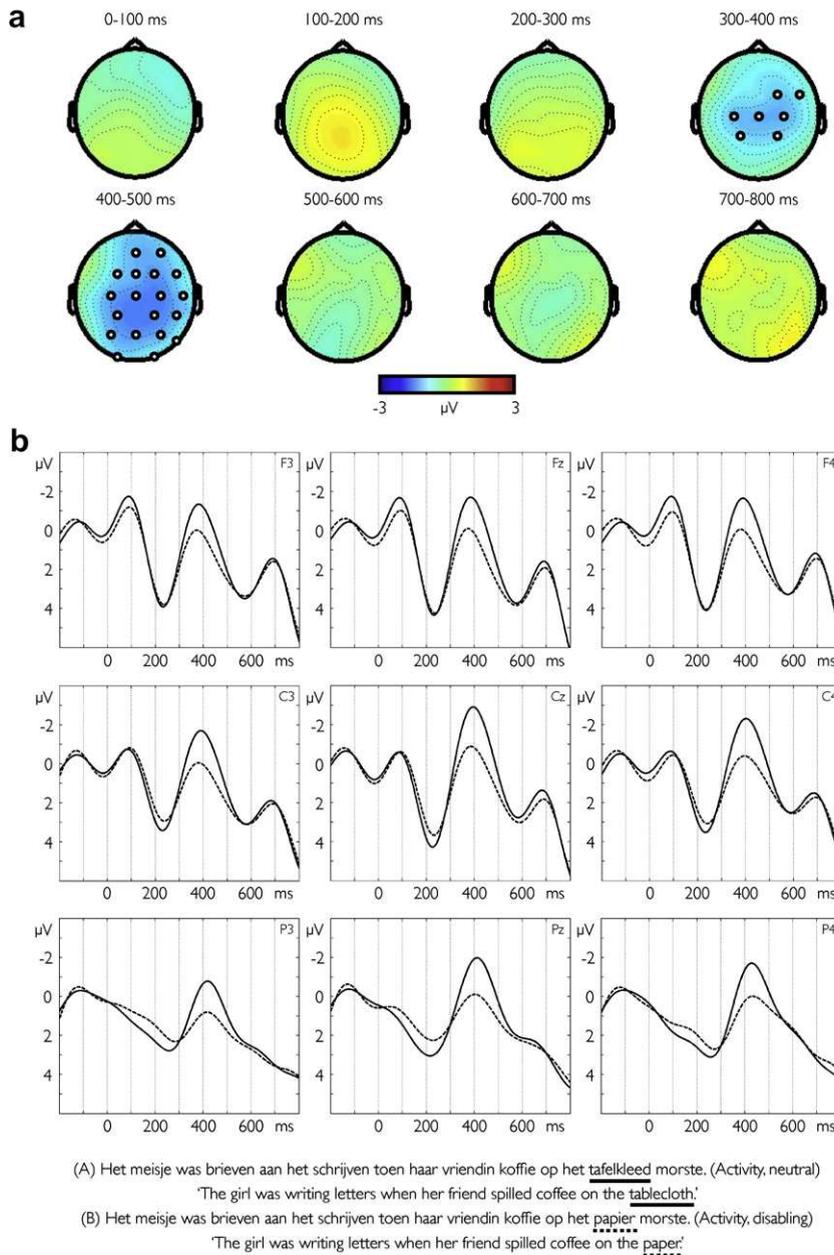


Fig. 2. Activities, noun. (a) Grand-average ($N = 24$) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled activities. Circles represent electrodes in a significant cluster. (b) Grand-average ($N = 24$) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled activities. Negative values are plotted upward.

at the noun, there is no difference between the N400 elicited by neutral and disabling clauses. Moreover, there is no difference between conditions in any of the remaining time bins (Table 3).

Fig. 5 displays the ERP topographies and waveforms elicited by *accomplishments*. No difference between disabling and neutral clauses was observed in either the

N1–P2 complex or the N400: no significant clusters between 0 and 400 ms were found (all contrasts, $P > 0.1$). While disabled and neutral *activities* do not result in any robust differential effect in later time bins (400–800 ms; Table 3 and Fig. 4), disabling verbs following *accomplishments* evoked larger negative shifts compared to neutral verbs (Table 3, Fig. 5). The effect

Table 3
Summary of cluster-based statistics for the ERP data

Time	Noun		Sentence-final verb	
	Activities	Accomplishments	Activities	Accomplishments
300–400 ms	$t(22) = -16.54$ $P = .026$	$t(22) = -60.11$ $P < .001$		
400–500 ms	$t(22) = -57.02$ $P < .001$	$t(22) = -78.69$ $P = .002$		$t(22) = -11.85$ $P = .034$
500–600 ms		$t(22) = -18.58$ $P = .022$		$t(22) = -71.09$ $P < .001$
600–700 ms				$t(22) = -39.16$ $P = .008$
700–800 ms				$t(22) = -16.92$ $P = .028$

Disabling and neutral clauses are compared at the noun and at the sentence-final verb, for activities and accomplishments, in bins of 100 ms starting from word onset. The first significant effects occurred at 300–400 ms. Empty cells denote the absence of significant clusters.

emerges at about 400 ms following the onset of sentence-final verb, lasts for approximately 400 ms, and is larger over the more anterior scalp sites, in particular of the left hemisphere. Based on its temporal profile and scalp distribution, we take this effect to be an instance of sustained anterior negativity (SAN). The magnitude of the SAN effect is correlated with the frequency of negative judgments in the response task ($r = -0.415$, $t(22) = -2.140$, $P = .043$; Fig. 6): the higher the number of negative responses, the larger the amplitude of the SAN.

No difference between the two aspectual classes was found. Cluster-based t -statistics comparing mean ERP amplitudes in activities and accomplishments, again corresponding to testing the main effect of Aspectual Class in a parametric model, produced no significant clusters between 0 and 800 ms from noun onset (all contrasts, $P > 0.1$).

Discussion

The ERP results reported above can be summarized as follows: the N400 elicited by nouns is larger in neutral than in disabling clauses, following both activities and accomplishments. This can be explained by the lower degree of semantic association with the preceding words ('writing', 'letter' or 'letters') of the noun in neutral clauses ('tablecloth') compared to the noun in disabling clauses ('paper'). On the basis of our processing model, we predicted that disabled accomplishments would induce a different ERP response at the sentence-final verb compared to neutral accomplishments. This corresponds to the difference between the recomputation and the extension of the initial discourse model (see "Minimal models" section). The effect was expected to be (i) absent in activities and (ii) correlated with the frequency with which participants inferred that the goal state was not attained. These predictions were borne out. Disabled

activities did not modulate ERPs at the verb. Disabled accomplishments evoked sustained anterior negativities (SANs). Moreover, a correlation of the SAN amplitude with the frequency of negative judgments was observed. Taken together, our results would seem to offer some support for the recomputation hypothesis. Below we address a few alternative explanations of the data and some related outstanding issues.

Alternative explanations and outstanding issues

Local integration

An alternative account of the data would relate the observed effect to difficulty in integrating the sentence-final verb into the 'local', clause-level context, rather than to suppressing a 'global', discourse-level inference. If this were correct, a modulation of the N400 should be expected, possibly correlated with differences in cloze probabilities. However, as reported above, cloze probabilities do not differ between conditions (see "Cloze probability test" section). Also, there was no difference in the N400s elicited by sentence-final verbs (see "Event-related brain potentials" section), which were also lexically identical across conditions. Following earlier work (Osterhout, 1997), we see the sustained anterior negativity as reflecting difficulty in constructing a discourse-level representation of disabled accomplishments. Supported by further experimental evidence, the recomputation hypothesis could provide a more explicit characterization of at least one instance of sentence-final 'wrap-up effects', in terms of restructuring the initial model.

Response frequency

Another alternative account would be based on the observation that sentences requiring a negative response (disabled accomplishments) are relatively less frequent than sentences requiring a positive one (activities and

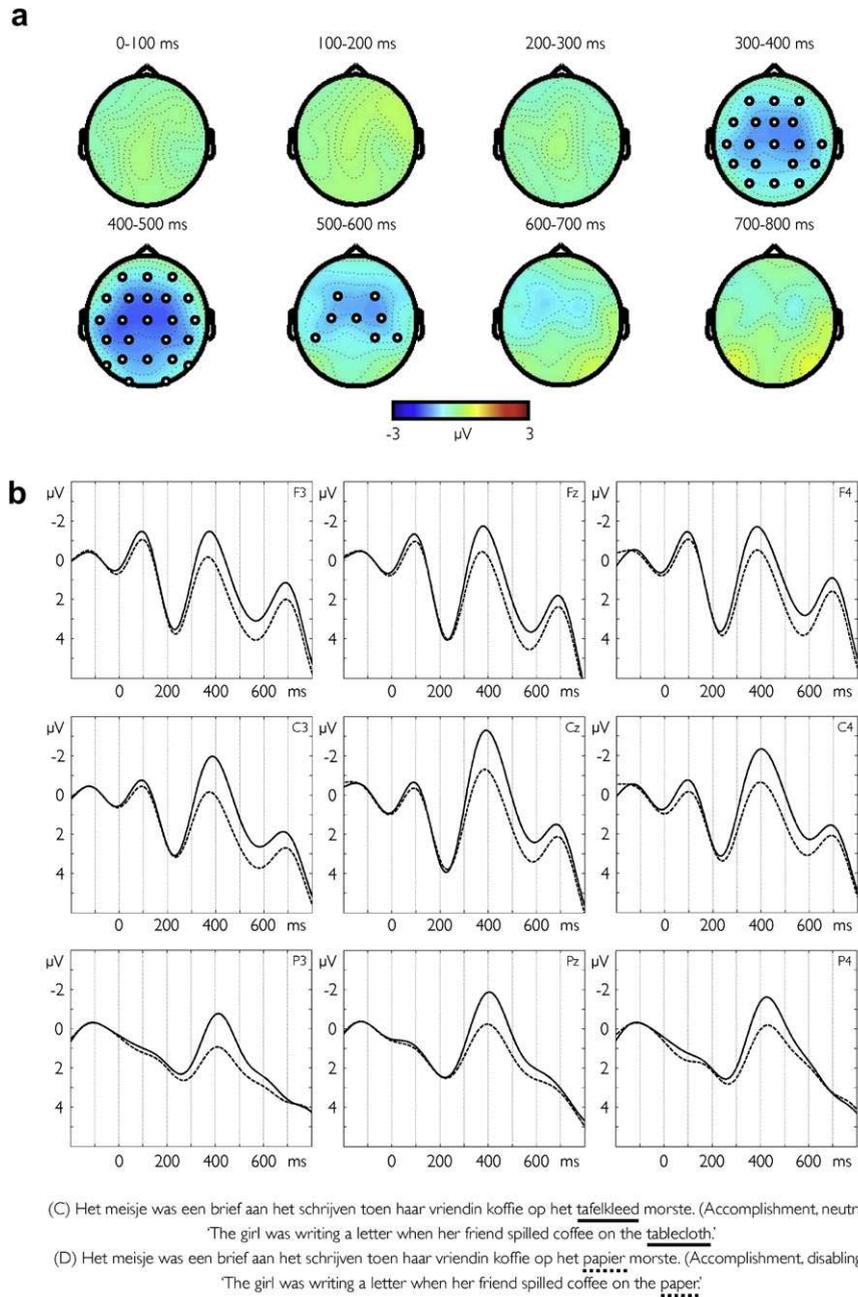


Fig. 3. Accomplishments, noun. (a) Grand-average ($N = 24$) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average ($N = 24$) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled accomplishments. Negative values are plotted upward.

neutral accomplishments), the projected ratios being respectively 1/4 and 3/4 (see “Behavioral data” section for the actual behavioral data). On this view, a modulation of the P3 component (Donchin, 1981; Ruchkin, Johnson, Cacoune, Ritter, & Hammer, 1990) might be expected, inversely correlated with the frequency of nega-

tive judgments to disabled accomplishments: the less frequent the negative responses, the larger the amplitude of the P3. However, in our experiment no P3 response was observed and, moreover, the correlation was rather the inverse: the more frequent the negative responses, the larger the amplitude of the sustained anterior negativity.

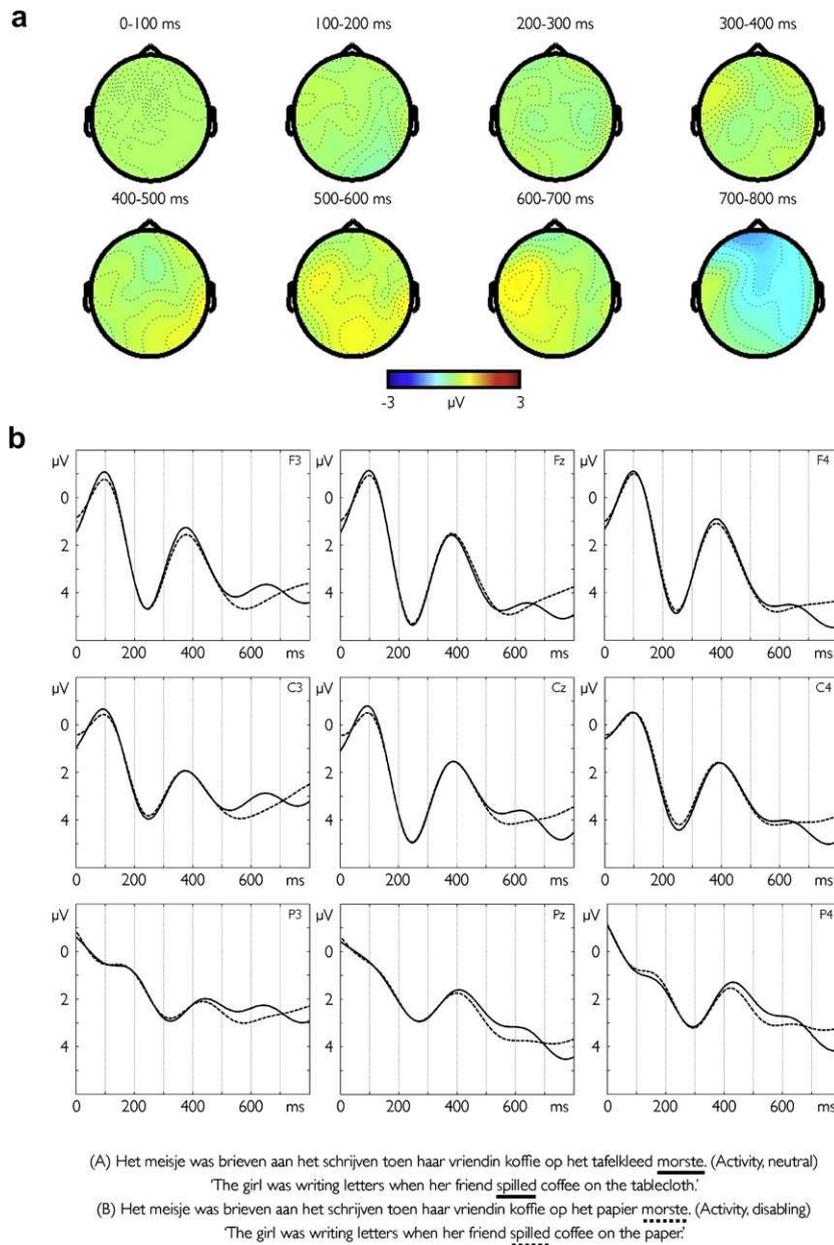


Fig. 4. Activities, sentence-final verb. (a) Grand-average ($N = 24$) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral activities. (b) Grand-average ($N = 24$) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled activities. Negative values are plotted upward.

Monotonicity and possible worlds semantics

An important issue is whether the observed sustained anterior negativity can be explained by a monotonic account of the progressive. Similarly, it may be asked whether the data reported here constitute compelling evidence for non-monotonicity and against monotonicity. One such alternative explanation can be formulated in possible worlds semantics (Dowty, 1979; Kripke, 1963).

In possible worlds semantics, the progressive denotes a stage of a process which, if it does not continue in the actual world, has chances of continuing in some other possible world (de Swart, 1998). The latter may be called ‘inertia worlds’, courses of events in which the process is not disturbed by external forces and is therefore brought to a successful end. In his analysis of the progressive, Dowty (1979) claimed that the following are equivalent:

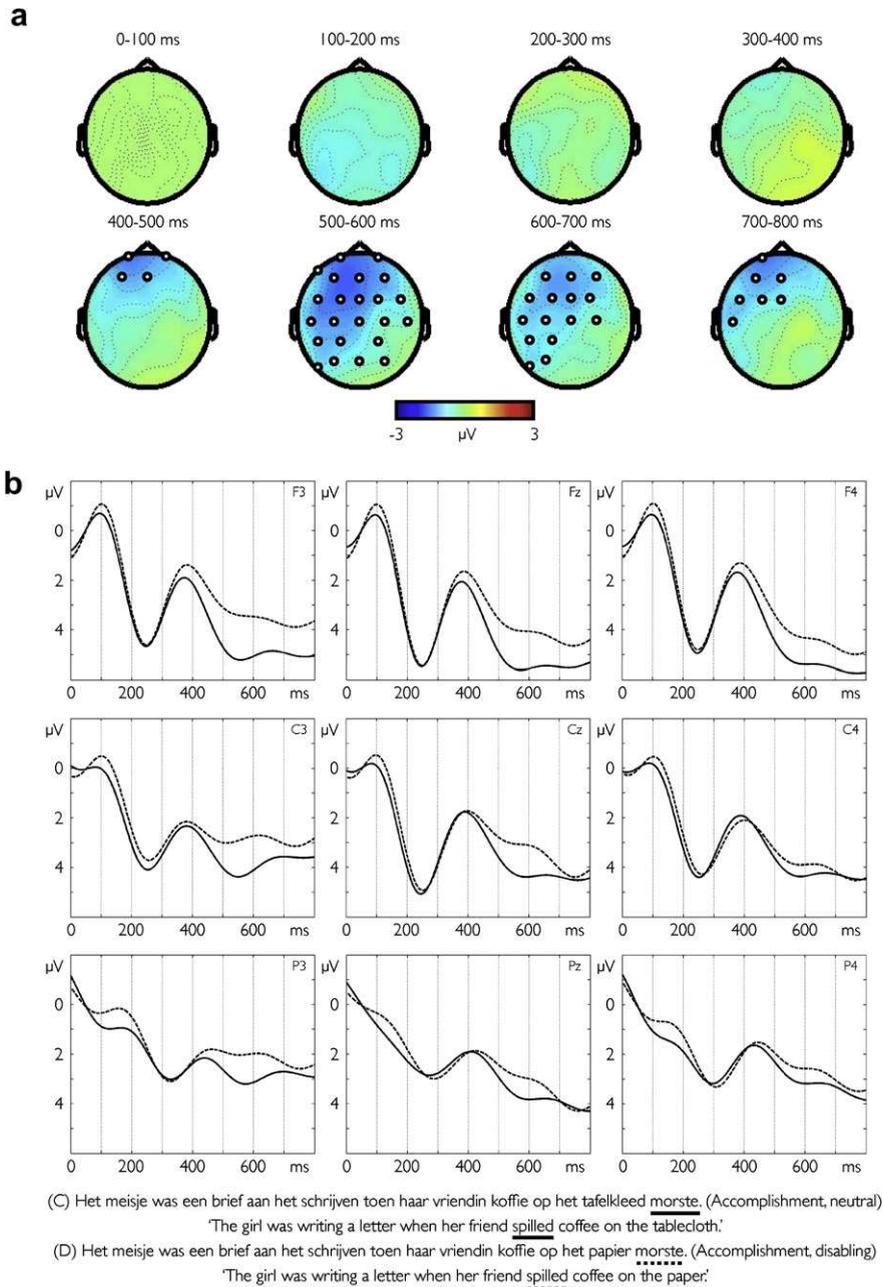


Fig. 5. Accomplishments, sentence-final verb. (a) Grand-average ($N = 24$) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average ($N = 24$) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled accomplishments. Negative values are plotted upward.

1. 'The girl is writing a letter' is true in the actual world;
2. 'The girl will have written a letter' is true in all so-called 'inertia worlds', worlds which are identical with the present world until 'now', but then continue in a way most compatible with the history of the world until 'now'.

These insights can be rendered into processing terms. Processing neutral accomplishments involves moving from the actual world, a snapshot of which is provided by the progressive clause, to some inertia world, in which the goal state is eventually attained (the behavioral data reported in "Behavioral data" section show

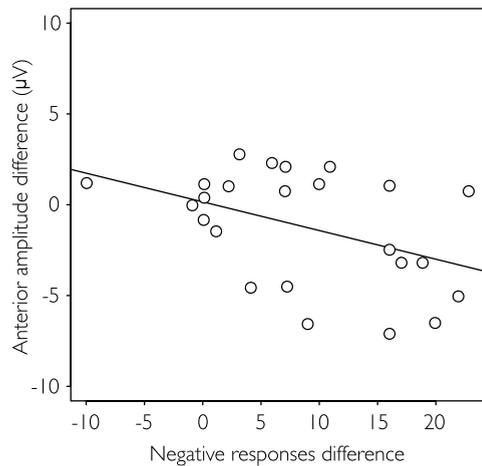


Fig. 6. Scatter plot displaying the correlation between the amplitude of the sustained anterior negativity elicited by disabled accomplishments and the frequency of negative responses ($r = -0.415$, $t(22) = -2.140$, $P = .043$). The mean difference of negative responses between disabled and neutral accomplishments is plotted on the abscissa. The mean amplitude difference at fronto-polar and frontal electrodes between disabled and neutral accomplishments in the 500- to 700-ms interval following the onset of the sentence-final verb is plotted on the ordinate.

that such an inference is drawn). By contrast, processing disabled accomplishments amounts to proceeding from the actual world to a relatively unexpected ‘non-inertia world’, in which the process is disrupted by some event, such as spilling coffee on the paper. Accessing a world in which the goal state is not attained may be surprising.² The sustained anterior negativity may then be construed as an index of surprise or some other equivalent notion. This account is monotonic, as for both neutral and disabled accomplishments interpretation involves shifting from the actual world to another accessible world. A simple extension of the initial model is performed in both cases.

This analysis is seemingly in conflict with the non-monotonic one. Still, there is no real opposition between the two as regards these data. In both accounts, an initial commitment to the occurrence of the goal state is made. In the non-monotonic approach this takes the form of a defeasible inference leading to a minimal model, whereas in the monotonic theory it is rather a prior, positive expectation concerning the attainment of the consequent state, and thus a lower probability

² This particular combination of possible worlds semantics and surprise was proposed by an anonymous reviewer of this paper, hence our choice of discussing it here. The same applies (for different reviewers) to the issues discussed in “Interruption and termination of activities” and “Goal states and underspecification” sections.

assigned to its failure. This commitment is necessary, because accessing a possible world in which the goal state is not satisfied can be surprising only if there is such a prior expectation. Furthermore, in both accounts a subsequent revision of the initial commitment is made. In the non-monotonic analysis this is a recomputation of the initial minimal model, whereas in the monotonic one it is a recomputation of the initial (low) probability associated with the possible world in which the goal state is not attained.

So there is recomputation and non-monotonicity built into this ‘monotonic’ theory too. The possible worlds account is monotonic as far as *models* are concerned (models are always extended and never recomputed), but *expectations* change non-monotonically (probabilities are recomputed). While *prima facie* opposed, the two accounts are in this respect equivalent. Our reasons for preferring a non-monotonic account, one in which *models* are recomputed, derive from *a priori* considerations. First, there exist forms of non-monotonic inference formally strongly related to the non-monotonic reasoning in the progressive which cannot be captured by Bayesian updates (Stenning & van Lambalgen, 2008a). Second, minimal models and non-monotonic inference can be implemented in neural networks (see “Recomputation in working memory models” section). Last, our account of the progressive is embedded into a larger non-monotonic framework (Stenning & van Lambalgen, 2008b; van Lambalgen & Hamm, 2004), covering other phenomena in reasoning and language processing in children, adults, as well as patients with ADHD (Attention Deficit Hyperactivity Disorder) (van Lambalgen et al., 2008) and ASD (Autistic Spectrum Disorder) (van Lambalgen & Smid, 2004).

We must note however that our study was designed to test a particular non-monotonic theory of the progressive, and not to discriminate between monotonic and non-monotonic accounts of the same phenomenon. The latter task would require, for one, a well-specified *entirely* monotonic theory—that is, one which does not involve recomputation of models, probability values or other processing parameters—and, moreover, a set of predictions in which the two proposals would actually differ. This is admittedly hard, apart from being beyond the scope of the research reported above. Hence the need to emphasize the direction along the theory-observation path which is relevant here: although it can be argued that our non-monotonic theory leads to predictions that are consistent with the observed sustained anterior negativity, it is clearly not the case that the data support *only* this particular theory.

Interruption and termination of activities

It may be argued that, compared to disabled accomplishments, disabled activities are inherently simpler

because they involve at most an *interruption* of the activity, for example writing letters in (B), which may be continued on some other paper sheets. Accomplishments might leave a more definite ERP trace because they lead to the *termination* of the activity, for example writing a letter in (D), which cannot be continued being there only a single sheet. On this view, the sustained anterior negativity would not be related to model recomputation (as opposed to monotonic extension), but to the termination (as opposed to the interruption) of the activity. Such an explanation follows from the seemingly plausible notion that computing a model in which the effects of a given event are more ‘catastrophic’ should also be more difficult. Here semantic theory comes to our rescue and suggests that such notion is in fact misguided.

One issue that plays a role here is a type/token distinction concerning the noun ‘letter’. In the token interpretation of ‘letter’ as referring to some particular scribbles on a particular piece of paper, there is indeed a difference between interruption and termination. However, on a type interpretation of ‘letter’ as referring to particular content which can be inscribed on any piece of paper, the activity and the accomplishment case seem comparable, in that in both cases the girl has to reach for a new piece of paper. On the type reading, one wouldn’t even expect a difference in behavioral responses. Nonetheless, since a behavioral difference was observed, it seems the token reading is what subjects adopt. On this assumption, it can be shown that, contrary to the alternative proposal, there is *more* computation going on in the interruption case compared to the termination case—if goal states are not taken into account; if they are, the pattern is reversed as implied by the recomputation hypothesis. It seems harder to compute a model in which an activity is first interrupted, then re-initiated, compared to computing a model in which the activity is just terminated (van Lambalgen & Hamm, 2004). The alternative account would predict a larger sustained anterior negativity for activities compared to accomplishments, which does not fit the experimental results. Also in this case, however, we are ready to acknowledge that a different model, in which terminations are shown to be *more* costly than interruptions, and in which goal states are not invoked to account for such processing cost, may explain the observed sustained anterior negativity.

Goal states and underspecification

The processing model adopted here implies that, as soon as an accomplishment in the past progressive is encountered, the system constructs a semantic representation in which the goal state is satisfied. Processing the clause ‘The girl was writing a letter’ amounts to computing a minimal model in which the writing activity leads to a complete letter, which is therefore part of the resulting discourse structure. As we have hypothesized, such

computation is defeasible, that is, the model can be recomputed if further discourse information implies that the goal state is not satisfiable, as in (6). One may ask whether the claim that the goal is part of a minimal model of the progressive clause is at all tenable. A seemingly more plausible account would assume that an *underspecified model*, in which it is left undecided whether the goal state is attained or not, is computed while the progressive clause is processed, and a decision is made only at the subordinate clause.

The main problem with an underspecification-based account is that, while it is true that the information provided by the progressive clause is insufficient for determining whether the goal was attained (which would motivate the construction of an underspecified model at that stage), it is *not* the case that sufficient information is contributed by the subordinate clause. While disabling clauses provide evidence that the activity was terminated, and thus license the inference that the goal was not attained, no evidence concerning the satisfaction of the goal state is derivable from neutral clauses. This is a consequence of the well-known ‘frame problem’ (McCarthy & Hayes, 1969), which implies that it is impossible to enumerate all the effects and non-effects of an event. For example, that ‘spilling coffee on the tablecloth does not affect the writing activity’ (if that is the case) is not stored in declarative memory, but must be inferred. This is an instance of ‘closed world reasoning’, which was described above (see “Minimal models, inference in the limit and recomputation” section). In a ‘closed world’, it is assumed that no obstacle to attaining the goal state occurred. Therefore, a letter was completed. The behavioral data reported above show that subjects draw this inference or, equivalently, they are more likely to give positive responses to neutral accomplishments. Processing models based on underspecification—or on parallel processing, for that matter—would have to explain why that very same conclusion (‘the girl wrote a letter’) is not drawn when the system is faced with the relevant input (the VP in the progressive), and is instead delayed until the end of the sentence, where critical information is nonetheless still missing. The hypothesis that the goal state inference is drawn when the input is given seems to be more consistent with the available evidence on immediacy and incrementality in discourse processing (Hagoort & van Berkum, 2007).

This line of reasoning speaks also to the issue of the potential influence of the primary response task on online interpretive processes. It can be argued that the system may have carried out a number of inferences online in order to facilitate a response when the probes were presented, but would have processed the same sentences in an underspecified manner if no response task was administered. The brain would therefore compute representations which are merely ‘good enough’ for the task at hand, striking a balance between efficiency and

cost minimization (Douglas & Martin, 2007; Ferreira, Ferraro, & Bailey, 2002; Ferreira & Patson, 2007). We grant that this is a possibility, which cannot be excluded based on either our data or our processing model. It can however be suggested that, although comprehension probes do not occur in actual language use, it is possible to imagine ‘language games’ in which hearers are required to make interpretive commitments and form a belief concerning the potential outcomes of a process described using the progressive. Our experiment may be taken as a laboratory study of such real world situations, but is not intended as a realistic account of all situations in which progressive constructions are uttered and understood. Further work is needed to investigate the influence of the response task on online ERP measures.

Recomputation in working memory networks

Minimal models can be regarded as the stable states of associated neural networks. It has been shown that recurrent networks can compute or approximate (depending on the expressiveness of the logical formalism) the semantic operators based on which minimal models are constructed (Hitzler, Hölldobler, & Seda, 2004; Stenning & van Lambalgen, 2005, 2008a). In this framework, recomputation can be modeled as the readjustment of connection strengths driven by a simple form of back-propagation called ‘perceptron learning’ (Rosenblatt, 1962). Computing a minimal model of the progressive clause will correspond to the network settling into one such ‘attractor’ or stable state. Further computation on the initial model brings the network from its initial stable state to another stable state, corresponding to the new minimal model. Importantly, there is a large difference in the overall pattern of network activity in disabled compared to neutral accomplishments. If the initial model is monotonically extended, as in the neutral case (4), a number of units will be activated which were previously silent, while the activation state of the remaining units, including those representing the goal state (the complete letter), will remain unaltered. But if the initial minimal model is recomputed upon encountering the subordinate clause in (6), units which were silent will be activated *and* the activation patterns across some units which were previously active will be readjusted. For instance, the units representing the goal state (the complete letter) will no longer be active. In the neural network this is achieved by successive applications of perceptron learning.

Even though in both cases the network processes the subordinate clause by settling into a new attractor state, the transition in the disabling case requires an extensive adjustment of the connection weights of the units representing the goal state. Recomputation thus results in a more costly state transition. It remains an open question

whether biologically plausible networks can also approximate the semantic operators which give rise to minimal models. Firing rate models, for instance, have been used to implement operations in connectionist networks (e.g. multilayer perceptrons) of the kind required by the construction of minimal models (de Kamps & van der Velde, 2002). Interestingly, recurrent excitation in firing-rate models can account for several aspects of persistent activity in prefrontal cortex neurons during working memory tasks (Durstewitz, Seamans, & Sejnowski, 2000). Recurrent networks thus suggest a plausible mechanistic link between recomputation and sustained anterior negativities, and in general between working memory processes and sustained anterior negativities (Felser et al., 2003; Fiebach et al., 2002; King & Kutas, 1995; Müller et al., 1997; Münte et al., 1998; Phillips et al., 2005; van Berkum et al., 1999, 2003).

As we noted in Introduction, a cognitive neuroscience of language needs to bridge the gap between psycholinguistic and formal models of specific aspects of language on the one hand, and the neural architecture underlying neurophysiological measures on the other hand. For a number of reasons (Poeppl & Embick, 2005) this is a daunting task, which we do not claim to have adequately solved. However, tentatively the following can be said. There is no indication or proof that the sustained anterior negativity is a language-specific ERP effect. Most likely, it reflects the recruitment of neurophysiological activity that might be generated in prefrontal cortex, and is triggered by different cognitive operations which build upon working memory capacity. For this purpose, the prefrontal cortex is a plausible candidate from a neurobiological point of view. In the light of our model, the sustained anterior negativity is taken to index the recomputation following the blocking of the goal state in accomplishments, and the recruitment of working memory resources required for this recomputation. In other cases, the demand might be triggered by different cognitive operations, as in the work by Münte et al. (1998). In general, what we seem to obtain with ERPs, is a many-to-one mapping from cognitive models to neuronal implementation. This however in no way invalidates our interpretation, which is based on combined constraints from the cognitive and neuronal levels of analysis.

The research presented in this paper extends the range of phenomena to which ERPs can be applied, by testing a processing hypothesis suggested by a formal semantics of tense and aspect. Our results raise the possibility that the brain supports some form of non-monotonic recomputation to integrate information which invalidates previously held assumptions. It is a task for future research to provide more stringent tests of non-monotonic as opposed to monotonic models of semantic processing and cognitive update more generally.

Acknowledgments

We thank Oliver Bott, Travis Choma, Bart Geurts, Fritz Hamm, Karl Magnus Petersson, Jos van Berkum, Roel Willems and three anonymous reviewers for many useful comments on earlier versions of the paper. We are grateful to the Netherlands Organization for Scientific Research (NWO) for support under Grant 051.04.040.

References

- Baayen, R., Piepenbrock, R., & Gulikers, L., (1996). *CELEX2*. Philadelphia: Linguistic data consortium.
- Baggio, G., & van Lambalgen, M. (2007). The processing consequences of the imperfective paradox. *Journal of Semantics*, 24, 307–330.
- Brown, C., & Hagoort, P. (1999). The cognitive neuroscience of language: Challenges and future directions. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 3–12). New York: Oxford University Press.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A, 639–663.
- Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.
- de Kamps, M., & van der Velde, R. (2002). Implementation of multilayer perceptron networks by populations of spiking neurons using rate coding. *Neurocomputing*, 353–358.
- de Swart, H. (1998). Aspect shift and coercion. *Natural Language and Linguistic Theory*, 16, 347–385.
- Donchin, E. (1981). Surprise!...surprise? *Psychophysiology*, 18, 493–513.
- Douglas, R., & Martin, K. (2007). Mapping the matrix: The ways of the neocortex. *Neuron*, 56, 226–238.
- Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Durstewitz, D., Seamans, J., & Sejnowski, T. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, 3, 1184–1191.
- Felser, C., Clahsen, H., & Münte, T. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87, 345–354.
- Ferreira, F., Ferraro, V., & Bailey, K. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Ferreira, R., & Patson, N. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83.
- Fiebach, C., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language*, 47, 250–272.
- Hagoort, P. (2006). The memory, unification, and control (MUC) model of language. In A. Meyer, L. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 243–270). Philadelphia: Psychology Press.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–81). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London, B*, 362, 801–811.
- Hamm, F., Kamp, H., & van Lambalgen, M. (2006). There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 22, 1–40.
- Hitzler, P., Hölldobler, S., & Seda, A. (2004). Logic programs and connectionist networks. *Journal of Applied Logic*, 3, 245–272.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language*. Amsterdam: Mathematical Centre.
- King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kripke, S. (1963). Semantical considerations in modal logic. *Acta Philosophica Fennica*, 16, 83–94.
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41, 142–151.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG and MEG data. *Journal of Neuroscience Methods*, 164, 177–190.
- McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4, 463–502.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition*, 23, 477–494.
- Müller, H., King, J., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193–203.
- Münte, T., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–73.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, 59, 494–522.
- Peregrin, J. (Ed.). (2003). *Meaning: The dynamic turn*. Oxford: Elsevier.
- Phillips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428.
- Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.),

- Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–118). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenblatt, R. (1962). *Principles of neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. New York: Spartan Books.
- Rothstein, S. (2004). *Structuring events*. Oxford: Blackwell.
- Ruchkin, D., Johnson, R. J., Cacoune, H., Ritter, W., & Hammer, M. (1990). Multiple sources of P3b associated with different types of information. *Psychophysiology*, *27*, 157–175.
- Steedman, M. (1997). Temporality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 895–938). Amsterdam: Elsevier.
- Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*, 919–960.
- Stenning, K., & van Lambalgen, M. (2008a). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.
- Stenning, K., & van Lambalgen, M. (2008b). The logical response to a noisy world. In M. Oaksford (Ed.), *The psychology of conditionals*. Oxford: Oxford University Press.
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*, 477–488.
- van Berkum, J., Brown, C., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- van Berkum, J., Brown, C., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*, 235–248.
- van Berkum, J., Hagoort, P., & Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- van Berkum, J., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701–718.
- van Lambalgen, M., & Hamm, R. (2004). *The proper treatment of events*. Oxford: Blackwell.
- van Lambalgen, M., & Smid, H. (2004). Reasoning patterns in autism: Rules and exceptions. In L. Perez Miranda & J. Larrazabal (Eds.), *Proceedings of the eighth international colloquium on cognitive science Donostia—San Sebastian*. Dordrecht: Kluwer.
- van Lambalgen, M., C. van Kruistum, & M. Parriger, (in press). Discourse processing in attention-deficit hyperactivity disorder (ADHD). *Journal of Logic, Language, and Information*.