LANGUAGE, LINGUISTICS AND COGNITION

GIOSUÈ BAGGIO, MICHIEL VAN LAMBALGEN, AND PETER HAGOORT

Author affiliations:

GIOSUÈ BAGGIO

F.C. Donders Centre for Cognitive Neuroimaging

Radboud University Nijmegen

MICHIEL 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

ABSTRACT. This chapter provides a partial overview of some currently debated issues in the cognitive science of language. We distinguish two families of problems, which we refer to as 'language and cognition' and 'linguistics and cognition'. Under the first heading we present and discuss the hypothesis that language, in particular the semantics of tense and aspect, is grounded in the planning system. We emphasize the role of non-monotonic inference during language comprehension. We look at the converse issue of the role of linguistic interpretation in reasoning tasks. Under the second heading we investigate the two foremost assumptions of current linguistic methodology, namely intuitions as the only adequate empirical basis of theories of meaning and grammar and the competence-performance distinction, arguing that these are among the heaviest burdens for a truly comprehensive approach to language. Marr's three-level scheme is proposed as an alternative methodological framework, which we apply in a review of two ERP studies on semantic processing, to the 'binding problem' for language, and in a conclusive set of remarks on relating theories in the cognitive science of language.

1. INTRODUCTION

This chapter provides a partial overview of some current issues in the cognitive science of language. The landscape that we shall try to render is vast and scattered, therefore it may be useful to start sketching the main lines using broad strokes. We make a distinction between two families of problems, which we refer to as 'language and cognition' and 'linguistics and cognition'.

Under the first heading – language and cognition – we will consider two problems. The first is which cognitive systems or functions support language. A number of 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 argued that meanings of concrete nouns stored in memory include stereotyped visual-geometric descriptions of objects (54). Analogously, representations of action verbs might embrace some aspects of the relevant motor programs (41; 88; 87; 89). It has been suggested that even the basic building blocks of natural language semantics like the predicate-argument structure might derive from the functional and anatomical organization of the visual and auditory brain systems (53). Language might also exploit the resources of higher cognitive domains such as action planning. Linearization as involved in executing a set of actions leading to a goal seems to be required by language production too, where the system has to transform a not necessarily linear conceptual and syntactic representation into speech.

Recent proposals have suggested that planning might as well support the representation and computation of syntactic and semantic structures: the recursive organization of plans would supply a basis for combinatory operations in grammar (99), while the goal-directed nature of planned actions would constrain cognitive constructions of time, causality and events, with some important consequences for the semantics of tense and aspect (99; 114). Semantic processing, involving the recomputation of discourse models (2) when additional information is provided, might be based on the non-monotonic logic underlying planning, and more generally executive control, allowing for action sequences to be readjusted when obstacles are encountered along the way (114; 101; 50). We will return to the issues of tense, aspect and the logic of discourse processing in several occasions in this chapter.

The second problem, which has captured the interests of some philosophers, is which cognitive functions are triggered or enhanced 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 phylogeny and ontogeny (13; 79). Particularly relevant for our purposes is the issue of the relations between interpretation and reasoning. Some have regarded non domain-specific thought and reasoning as the most prominent of the cognitive skills subserved by language (11; 12). This suggestion is often implicit in the tradition of logical approaches to language since Boole[1] and bears some resemblance to the psycholinguistic notion that reasoning follows and builds upon interpretation (57). On this view, inference elaborates and enriches propositional meanings (96) by establishing a number of conceptual links with other types of representation, such as world knowledge and action schemes. Linguistic interpretation would then subserve reasoning, but not vice versa. Others have seen the relation between inference and interpretation as reversed, if not based on an interaction (101; 102). Human communication is thus 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 (100). We will touch upon this topic in section 2.3.

Under the second heading – linguistics and cognition – we include two broad issues. The

[1]In 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." (5, Ch. 2, p. 24)

first is whether the structures and rules described by linguists are 'cognitively real'. There exist several opinions in this regard, occupying different positions on the spectrum between mentalism and anti-mentalism. At one extreme is cognitive linguistics (24), endorsing both theoretical and methodological mentalism. As for theoretical mentalism, cognitive linguists view linguistic structures as representations entertaining causal and conceptual relations with other mental entities. Some proponents of cognitive approaches to language have also called for a revision of linguistic methodology in order to stimulate some interaction with psycholinguistics and cognitive neuroscience (55). At the opposite side of the spectrum lies formal semantic theory which, inspired by Frege's anti-psychologistic stance on meaning and thought (9), rejects both theoretical and methodological mentalism. Semanticists in the tradition of Frege and Montague usually see linguistics as separate from the study of the social and cognitive aspects of language (63). As regards theoretical mentalism, Frege famously considered meanings as objective entities, not to be confused with the subjective representations associated to linguistic expressions.[2] Between the two poles is Chomsky's (17) theoretical mentalism, viewing syntactic rules as ultimately residing in the mind-brain of speakers. However, his commitment to the cognitive reality of grammar does not imply a revision of linguistic methodology, which is on the contrary maintained in its traditional form as based on native speakers' intuitions and the competence-performance distinction. In section 3 we argue that these are among the heaviest burdens for a truly comprehensive approach to language.

The second problem, not independent of the first, is whether data from experiments on language acquisition, comprehension and production have any bearing on linguistic theory. On this point there is no consensus among linguists. For instance, the distinction between competence and performance has often been used to secure linguistics from experimental

[2]In *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." (34, pp. 24-25)

evidence of various sort (8), while intuitive judgments of native speakers were regarded as the only type of data relevant for the theory (17). However, some commentators have granted that at least behavioral data should be allowed to inform theories of competence, for example when the linguist is studying a language which is not her own native language (66). Others, most notably Jackendoff (55), have proposed frameworks in which theories of competence can be explicitly related to performance mechanisms. Although these models account for how competence constrains performance, they overlook the possibility that the reverse is also true. In fact, at least some aspects of competence might be the outcome of an evolutionary process leading to an adaptation of the system to language use, that is to performance (86).[3] While generative grammar and formal semantics have often regarded theories of competence as shielded from empirical data of the kind customarily provided by cognitive psychology and neuroscience, a slightly more liberal attitude has been adopted by psycholinguists and cognitive brain scientists. In several studies linguistic hypotheses have been discussed in the light of experimental data and in some cases have been explicitly tested (see for instance (10; 28; 36; 72; 73)). Despite these fine attempts, which undeniably constitute a progress, frameworks accommodating linguistic theories, language processing models and low-level mechanistic descriptions are lacking since the publication of Chomsky's *Aspects of the Theory of Syntax* (17).

The aim of this chapter is to remove a few obstacles on the road and take some tentative steps in that direction. Although we will examine in detail several linguistic and cognitive phenomena, our main concern remains methodology: the problem of the relations between theories, as well as between empirical domains, in the cognitive science of language is of paramount importance for our purposes and will be extensively discussed in sections 3, 4 and 5. We start by introducing the test case on which a large portion of our arguments will be based: the semantics of tense and aspect.

2. LANGUAGE AND COGNITION

[3]The reference to evolution here may suggest that we take biological inheritance to be the key factor in the process leading to a relatively stable system of competence rules. However, it should be emphasized that cultural inheritance seems to play at least an equally important role. For convincing cases in favor of this point, the reader is referred to the works of Tomasello(106), Cavalli-Sforza & Feldman (13) and Nowak et al. (79). We return to the issue of adaptation to performance in section 3.3.

2.1. **Cognitive substrates of tense and aspect.** We take it as the essential purpose of tense and aspect to facilitate the computation of event structure as described in a narrative. One implication of this characterization is that, contrary to what generative and formal semantic approaches 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 (21; 22). Tense and aspect, however, cannot by themselves completely determine event structure, and must recruit world knowledge. Examples (1a-c) below will make clear what we have in mind.

French has several past tenses (Passé Simple, Imparfait, Passé Composé), which differ in their aspectual contributions. The following mini-discourses in French all consist of one sentence in the Imparfait and one in the Passé Simple; however, the structure of the set of events differs in each case:

(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 example, the Imp-sentence describes the background against which the event described by the PS-sentence occurs. In the second example, the event described by the PS terminates the event described by the Imp, whereas in the third example the relation is rather one of initiation. These discourses indicate 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, but may be consciously recruited if the automatic processing leaves the event structure still underdetermined. As we will argue in section 3, it is the task of cognitive science to determine what this algorithm is and how it is actually implemented. The prominent role of causal relationships highlighted by (1a-c) 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.

Here we present several converging lines of evidence which lend some plausibility to this conjecture. Firstly, 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 is accompanied by 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 may involve a good deal of deliberation. More precisely, planning consists in

> the construction of a *sequence*[4] 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 origins of language. Even if it is granted that some non-human 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.[5] 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, that is the utterance; as we have seen, planning also involves linearization. In the next step of the argument, the recursive structure of syntax is linked to the recursive structure (hierarchical organization) of plans (39; 99). That is, planning is used to explain both the continuity with non-human primates and the divergence. Both humans and non-human primates engage in planning.

[4]More complex plans are possible which may involve overlapping actions.

[5]The issue of recursiveness in animal communication is too involved to be addressed here. For examples, see (30) for a study on grammar discrimination in cotton-top tamarins (*Saguinus oedipus*) and (35) for a similar experiment with European starlings (*Sturnus vulgaris*). For a discussion see (82).

Primates are adept at planning, at least for time spans not exceeding 48 hours, as is known since Kohler's 1925 observations (58). It has also been attested in squirrel monkeys (*Cebus apella*), in experiments by McGonigle, Chalmers & Dickinson (71).

A more direct link between language and planning was investigated experimentally by Trabasso & Stein (108) 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*. They defend the thesis that "the plan unites the past (a desired state) with the present (an attempt) and the future (the attainment of that state)" (108, p. 322) and "[c]ausality and planning provide the medium through which the past is glued to the present and future" (108, p. 347). Trabasso & 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 jar.[6] The drawings depict various failed attempts, until the boy finds his frog by accident. The purpose of the study is to determine what linguistic devices, in particular temporal expressions, children use to narrate the story as a function of age. The authors provide some protocols which show a child of age 3 narrating the story in a 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 the 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 relation 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 is capable of including 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 following quote suggests that there must be a *gradual* development of these capabilities:

[6] This is a classic experimental paradigm for investigating the acquisition of temporal notions in children. See (3) for methods, results and, last but not least, the frog pictures themselves.

> 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. (108, 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 expressions.

What is interesting for our purposes is that the ingredients that jointly enable planning have a prominent role to play in the construction of discourse models. Take for instance causality, shown to be involved in the interpretation of (1a-c). Planning essentially requires knowledge of the causal effects of actions as well as of the causal effects of possible events in the world. Accordingly, the planning capacity must have devised ways of retrieving such knowledge from memory. Planning also essentially involves ordering actions with respect to each other and to events occurring in the world which are not dependent on the agent. Furthermore, the resulting structure must be held in memory at least until the desired goal is attained. The reader can easily envisage this by considering the planning steps leading to a pile of pancakes – for instance, 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 processing and planning is that both are non-monotonic. When we plan, deliberately or automatically, we do so in virtue of our best guess about the world in which we 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 model is non-monotonic (2). For instance, the reader who sees

(2a) is likely to infer (that is, to read off from the discourse model) that Bill is no longer a member, but that implicature can easily be canceled, as in (2b):

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

b. 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 which 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. Hence our main hypothesis can be reformulated as follows: several components of executive function are involved in comprehension and production of tense and aspect. A corollary is that failures of executive function can show up in deviant use of tense and aspect and in impairments in processing temporal discourse, for instance in ASD (Autistic Spectrum Disorder), ADHD (Attention Deficit Hyperactivity Disorder, see section 2.2.2) and schizophrenia.

The link between planning and temporal semantics 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:

1. location of event in time;
2. meshing it with other events.

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

(3) Max fell. John pushed him.

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

$$\text{`update discourse with past event } e_1 = \mathtt{fall}(m) \text{ and fit } e_1 \text{ in context'}$$

$$\text{`update discourse with past event } e_2 = \mathtt{push}(j,m) \text{ and fit } e_2 \text{ in context'}$$

Planning must determine the order of e_1 and 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 detailed picture of what goes on in such computations, we have to introduce some notation, borrowed from the Event Calculus (114), which will also be useful for our discussion of experiments on language processing and the 'binding problem' later on in this chapter. We make a distinction between *events* (denoted $e, e', ..., e_0, ...$) and *processes* or *fluents* (denoted $f, f', ..., f_0, ...$). We say that events *occur* or *happen* at a particular time, and represent this 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 one: $Terminates(e, f, t)$. We will use these predicates to mean the causal relation only; it is not implied that e actually occurs. Finally, a useful 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 shall see a glimpse below.[7] With this notation, and using $?\varphi(x)$ `succeeds` to abbreviate: 'make it the case in our discourse model that $\varphi(x)$',[8] we can write the two update instructions involved in comprehending the discourse as:

(4) $?Happens(e_1,t), t < now, Happens(e',t')$ `succeeds`

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

Here e' and e'' are variables for event types in the discourse context which have to be found out by substitution or, more precisely, *unification*. These two update instructions have to be executed so that $e'' = e_1$ and $s < t''$. If 'Max fell' is the first sentence of the discourse, we may disregard e'.[9] In order 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:

(6) $HoldsAt(f,t) \rightarrow Happens(e_1,t)$

(7) $Initiates(e_2, f, s)$

---

[7] We share the concerns of logical positivists and some contemporary philosophers of science, most notably Suppes (103; 104), to the effect that, for non-axiomatized theories, it remains unspecified which structures are represented by the theory and which consequences can be derived from its principles. The present theory of tense and aspect is fully axiomatized and is equipped with a meta-theory (114).

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

[9] Here we regard context as provided by the preceding discourse, although one may conceive of 'forward-looking' notions of context as well.

The system processing the discourse will first satisfy the update request corresponding to 'Max fell' by locating the event e_1 in the past of the moment of speech. The second sentence, 'John pushed him', is represented by the request (5) which contains the variable e''. The system will try to satisfy the goal by reducing it using relevant causal knowledge. Applying (6) and unifying[10] $e'' = e_1 = \mathit{fall}(m)$, the second request (5) is reduced to:

(8) $?\mathit{Happens}(e_2, s), s < \mathit{now}, \mathit{Happens}(e_1, t''), \mathit{HoldsAt}(f, t'')$ `succeeds`

The system now applies a general causal principle, 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 occurrences of events. Inertia can be formulated as the following axiom:

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

Using this axiom, the request (8) is further reduced to:

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

Using (7) and unifying $e = e_2 = \mathit{push}(j, m)$ and $s = t$ we reduce this request to:

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

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

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

Since we have no positive information to this effect, we may assume $\neg \mathit{Clipped}(s, f, t'')$. This form of argument is also 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, as it allows one to discount events which are logically possible but in practice irrelevant. The final update request is thus:

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

---

[10] This form of unification will be important in our discussion of the 'binding problem' for language.

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.

Just as plans may have to be revised in mid-execution (for instance, if it turns out there is not sufficient oil to produce the projected number of pancakes), discourse models may have to be recomputed when additional information is provided (2). We will elaborate our 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:

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

One obvious interpretation is that now $e_2 = push(j,m)$ comes after $e_1 = fall(m)$. This is the result of a recomputation, since after the first 'him' the hearer may have inferred that e_2 precedes e_1. Let us give an informal sketch of this recomputation. 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:

(15) $?Happens(e_2,s), s < now, HoldsAt(dead(m),s), Happens(e'',t'')$ `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 denoted by f), Max is still *alive*. Unifying $e'' = e_1$ and applying property (6), the request reduces to finding instants s, t'' such that:

(16) $?Happens(e_2,s), s < now, HoldsAt(dead(m),s), HoldsAt(alive(m),t''), Happens(e_1,t'')$ `succeeds`

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

In summary, what we have outlined here is a universal computational mechanism for determining event structure based on planning. Temporal expressions (not only tense and aspect, but also temporal connectives as will be seen in section 3.5.3) are hypothesized to determine requests to be satisfied by an update of the current discourse model. Processing these requests involves unification, search through semantic memory, as well as setting up temporary structures in working memory. These computations might leave characteristic traces on neurophysiological responses such as event-related potentials (or ERPs, see 3.5.1), a hypothesis to which we will return later.

2.2.1. *Computing event structures for (PS, Imp) combinations.* Similar arguments apply to 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 and the foregrounded event ('taking off one's sweater') is located within 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 time *t* before *now*. By the principle of inertia, the state *hot* must either hold initially (at the beginning of the narrative) or have been initiated. The latter requires the occurrence of an initiating event, which is however not given by the discourse. Therefore, *hot* holds initially. Similarly, no terminating event is mentioned, so *hot* extends indefinitely, and it follows that the event described by the second sentence must be positioned inside *hot*.

The second example dates back to the bygone days when speeding cars were stopped by the police instead of being photographed:

(1) b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)
Jean got a ticket. He was driving too fast.

It is given that the event of getting a ticket occurred sometime in the past, and it is also given that the fluent *speeding* was true sometime 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 does not initiate *speeding*. Because this is the only event mentioned, *speeding* holds from the beginning of discourse, and is not reinitiated once it has been terminated.

In the third example, 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, therefore the light remains on. It also follows that the light must be off for some time prior to being switched on and 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; as a consequence, it is *this* light which is too bright.

2.2.2. *Deviant verb tenses and ADHD.* The preceding considerations can be applied to those 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 execution – in short, this ability requires executive function as defined above (90).

The theory presented above 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 (37). Recall that update requests, that is the goals to be satisfied, corresponding to a VP's tense and aspect, consist of two components:

1. location of an event in time;
2. meshing the event with other events.

If a child has trouble maintaining a goal in working memory, this may lead to a simplified representation of that goal. In the case of verb tenses the most probable simplification is of '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 summarized in the following quote:

> 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. (90, p. 141)

This seems to suggest that there is indeed a connnection between language comprehension and production, and processing of goals.

2.3. **Why language matters to cognitive science.** Most 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 there 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 suggest, 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 (see (102) for a discussion). We believe that these views betray a lack of awareness of the tremendous variation in language understanding and especially language use.

2.3.1. *Reasoning in illiterate subjects.* A fundamental observation concerning reasoning in illiterate subjects is their obstinate refusal to draw conclusions from premisses supplied by the experimenter, as in Luria's (65) famous example:

> *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 on the basis of personal knowledge only, interfering with deductive inference.

A more 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 the 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 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 & Cole's (95) study of reasoning among the illiterate Kpelle tribe in Liberia. Here is a sample argument given to her subjects:

> All Kpelle men are rice farmers.
> Mr. Smith[11] is not a rice farmer.
> Is Mr. Smith a Kpelle man?

Consider the following dialogue with a subject:

> *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.

[11]'Mr. Smith' is not a possible Kpelle name.

On the plausible assumption that in the dialogue 'if' means 'if and only if' (102), 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 concluded that his subjects' thinking is limited to the concrete and the situational, that of which they have personal knowledge, while Scribner tended to the conclusion that subjects fail because they don't understand the problem or what kind of 'language 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 carry the hallmarks of moral scruples about bearing witness on the basis of hearsay.

We may note here that the experimenter expects 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 comprehension, it may very well be that this requires understanding of a particular discourse genre, roughly that of an exam question, 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! This makes inference from 'given premisses' a non-starter. Here is another example (65):

> *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 regard it as the construction of a discourse model of the premisses using world knowledge ('cold and damp implies poor soil') and solving the task by describing the model, in line with the studies on sentence memory by Bransford, Barclay & Franks (6).

2.3.2. *Rational reasoning and the material implication.* No other experiment in the psychology of reasoning has generated more controversy, and also more theories of reasoning, than the Wason Selection Task (118) (see Figure 1). The experiment has been replicated many times. The reader, who may have 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:

> 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:**

A	K	4	7

FIGURE 1. Wason's selection task

If one formulates the rule 'If there is a vowel on one side, then there is an even number on the other side' as the implication $p \rightarrow q$, then the observed pattern of results is as given in Table 1. The modal response (around half of the undergraduate subjects) is to turn A and 4. Very few 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 Wason to attribute 'irrationality' to his subjects, and others to deny that formal logic is relevant to actual human reasoning (see (102) for a review).

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 as the material implication. But nothing in the experimental material forces this reading, and the experimenter must assume that subjects come 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 can often be interpreted as exception-tolerant: 'If a patient has pneumonia, give him anti-biotics' is true, even if those patients with pneumonia who are allergic to anti-biotics must on no account get them.[12] Indeed, one finds subjects who cannot do the task as intended because they view the rule as allowing exceptions:[13]

> *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 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 participants indicates that they have problems grasping the notions of truth ('Turning A and 7 would at most show me the rule is not false; but why is it true?') and falsity, many interpreting '$p \rightarrow q$ is false' as meaning $p \rightarrow \neg q$, in which case the choice of the p card only is 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

[12]This observation suggests an alternative interpretation of the results of Luria (65). For 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 which the subject quite appropriately denies to possess.

[13]From an experiment reported in (102, Ch. 3), in which subjects are shown real cards and engage in conversation about their choices with the experimenter.

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

TABLE 1. Typical scores in the selection task

'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 methodological dilemma.** It is often 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* (17) in 1965. This is certainly true in view of Chomsky's theoretical mentalism, conceiving of linguistics as dealing ultimately with a system of representations and rules in the speaker's mind-brain. However, although this form of 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 opposite effect, that is separating theories of meaning and grammar from language processing models. Many commentators would agree that the contacts between linguistics and cognitive psychology have not been as deep and systematic as they could have been, had various obstacles been removed (55). What seems more difficult to appreciate is the existence of such inner tension in the very foundation of generative grammar and the inhibiting effect it had on the growth of linguistic theory within cognitive psychology and neuroscience. Before we move on to the more specific issues concerning linguistics and cognition discussed in this section, it may be worth recovering the terms of this 'dilemma' directly from Chomsky's text.[14]

One side of the dilemma 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-hearer's 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

---

[14] Chomsky has entertained different opinions on this issue throughout his career. Here we choose to focus on those expressed in *Aspects of the Theory of Syntax* (17) because these have probably been the most influential. Thus, we identify Chomsky with this particular text rather than with the actual linguist.

> underlying system of rules that has been mastered by the 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. (17, p. 4)

The task of the linguist is to provide an account of competence as based on performance data, that is on normalized records of linguistic behavior. Chomsky grants that performance data are essential to linguistic theorizing. The issue to be settled, which in fact lies at the heart of the dilemma, is exactly what counts as linguistic behavior or more precisely what kind of performance data can form the empirical basis of competence. Although generative linguists would contend that it was never a tenet of their research programme to admit data other than native speakers' intuitions, this is not what is suggested by Chomsky's remarks. On the contrary, he seems to admit 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. (17, p. 193)

The evidential basis of linguistics consists of introspective judgments *and* performance data, which Chomsky mentions 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 class of data types. The question is what should be considered performance data 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 his remarks on the limits of intuitions. In 1955 in *The Logical Structure of Linguistic Theory* (15) he wrote:

> 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'. (15, 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 and their constituents should be evaluated on the basis of controlled operational 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. (16, pp. 8-9)[15]

Controlled operational tests are thus necessary in order to overcome the difficulties arising from relying exclusively on native speakers' 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 dilemma 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

[15] The circularity which Chomsky is referring to here is also mentioned 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 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." (91, p. 392) Root (93) has argued that both Chomsky and Quine, despite the differences, were trying to define linguistic notions operationally, suggesting that Chomsky, at least in his early writings, allowed controlled tests of some sort to affect theories of competence.

> 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). (17, p. 18)

Experimental research based on controlled observation and statistical inference is seen as providing facts of no 'interest and importance' and rejected as ineffective for the purposes of the theory of competence. Interestingly, intuitions are treated as they were on a par with performance data. Not for long, however, since Chomsky a few paragraphs later takes an important 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. (17, pp. 19-20)

The door that seemed to be open for controlled experimentation is now closed. The range of data that could affect the theory of competence has been narrowed down to intuitions, and more specifically to those of the linguist. The task of experimental research, Chomsky says, is to develop tests that will ultimately align with introspective data. The convergence of 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 psychology are now regarded as separate enterprises, but the latter is also required – in order to meet the criterion of explanatory adequacy – to provide results that are consistent with the theory of competence as based on the linguist's intuitions.

We would like to note that, whatever the methodological choices made by linguists, this is an unacceptable requirement for cognitive psychology and neuroscience, which should instead be urged to provide solid evidence, no matter which theory or whose intuitions are eventually supported. If experimental research provides evidence which does not align with the introspective judgments of the linguist or other native speakers, then, following common practice in science, there is no other choice than to accept the results of the former and reject the latter. But what is perhaps most disturbing 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. (17, p. 20)

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

As we said at the beginning of the chapter, most generative linguists accept theoretical mentalism and reject methodological mentalism. However, this gives rise to the dilemma just described: the two views can be separated only at the price of a necessarily ambiguous attitude toward experimental data. A solution of the problem, one may argue, would then amount either to rejecting theoretical mentalism in order to maintain traditional linguistic methodology or to adopting a more receptive attitude toward the procedures and results of experimental psychology. In other words, the solution would be either to take up both theoretical and methodological mentalism or to reject both. Still, we believe, none of these would be sufficient. On the one hand, as we shall see in section 3.3, competence theories as they are usually conceived, namely lacking an explicit link with performance and more precisely with the notion of 'information processing', might be refractory to the results of cognitive psychology and neuroscience, whatever the linguist's attitude toward them. On the other hand, intuitions might still be an undesirable source of bias for the theory even if the evidential basis was enlarged so as to embrace 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 manner. 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 psycholinguist studying performance.

3.2. **The vagaries of intuition.** For many linguists, in particular generative grammarians and formal semanticists, the intuitions of native speakers constitute the empirical basis of theories of competence. But as we have just seen, the prominent place assigned to intuitions by current linguistic theory is at odds with the basic tenets of mentalism. If competence is a system of structures and rules realized in the speaker's mind-brain, and if behavior reflects the properties of such system, then a linguist constructing a theory of competence – and analogously a child learning a language – must solve an 'inverse problem', that is inferring the rules of competence from observable performance data.[16] In order to devise a solution, linguists might need to take into account a wide range of data: any reliable (physiological or behavioral) measure of performance should, at least in principle, be allowed to contribute to accounts of competence. In this regard, methodological mentalism seems a more suitable framework than standard linguistic methodology as based on intuitions.

The conflict with mentalism is however not the only problem raised by introspective judgments. Another source of concern is Chomsky's claim that intuitions are not only the starting point of linguistic theorizing, but also provide 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. (17, 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 inquiry, intuitions would still pose a number of serious methodological problems. It is not

[16] The solvability of such inverse problems has been discussed in formal learning theory, for instance by Gold (38).

just the role of intuitions in linguistic theorizing that has to be put under scrutiny, but also the very claim that intuitions offer a privileged window on tacit knowledge.

3.2.1. *Intuitions in linguistics.* Some commentators have questioned the claim that intuitions provide a transparent access to competence. For instance, Jackendoff has observed that the system of rules in the speaker's mind-brain is "deeply unconscious and largely unavailable to introspection" (56, p. 652). If this is correct, then one should see discrepancies between overt linguistic behavior, reflecting the 'unconscious' competence rules, and the intuitions or beliefs that speakers have about these rules. This idea has been substantiated by Labov (61), who collected linguistic evidence on a variety of cases in regional American English. One example is the positive use of 'anymore' in various 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:

(17) 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 (17) and reported a majority of negative responses when they were asked whether a sentence like (17) is acceptable, as well as surprisingly weak intuitions on what the expression signifies in their own dialect, 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 advanced. For instance, Marantz has recently emphasized that grammaticality is a technical term defined within linguistic theory (66): 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. But native speakers do not have intuitions of grammaticality in the technical sense, nor they have acquaintance of properties of strings as defined within a formal grammar. Furthermore, naïve language users might conflate into the notion of grammaticality different morpho-syntactic, semantic and pragmatic aspects of language, and they might do so in a way that is beyond control for the linguist. Similar observations would also apply to intuitive judgments of synonymy (or truth-conditions) as opposed to formal definitions within a semantic theory. Therefore, it may seem inappropriate to regard introspective evidence as being more than a starting point of linguistic inquiry.

As a way out, one might be tempted to argue that a caveat would only apply to naïve informants, and that intuitions are reliable when the judgments of trained linguists, free from pre-theoretical and confused notions of grammaticality, synonymy, and the like, are considered (see (25, pp. 499-500) for a recent defense of this claim). Relevant to this issue is a demonstration experiment by Levelt (62) in which the intuitions of twenty-four trained linguists were tested. Participants were presented with fourteen examples from their own field's literature, among which:

(18) a. No American, who was wise, remained in the country.
b. The giving of the lecture by the man who arrived yesterday assisted us.

None of the linguists judged the ungrammatical item (18a) as such and sixteen judged the grammatical item (18b) as ungrammatical. Overall, ungrammatical items had less chance of being judged ungrammatical than grammatical items. Levelt warns against taking these results too seriously, but he observes with some reason that "they are sufficiently disturbing to caution against present day uses of intuition" (62, p. 25).

We could go on providing other examples of the problems that might arise with the use of introspective reports in the analysis of specific natural language sentences. However, we would now like to take a different approach, considering an argument targeted at the very nature and scope of intuitions. The argument, which has been proposed and discussed by Hintikka (52), starts off with the observation that our intuitions of grammaticality and synonymy always relate to particular sentences or tokens, and not entire classes of items or the common syntactic or semantic 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. (52, p. 137-138)

Against Hintikka, one may argue that also paradigmatic variation is a proper object of intuition. The linguist would then be able to generalize over the properties of linguistic structures by constructing a paradigmatic set of sentences exhibiting those properties. This claim however can be countered with the observation that the supposed 'intuitions' about paradigmatic cases are more similar to theory-laden hypotheses than to the introspective judgments of naïve informants. For the linguist, in order to construct such paradigmatic items, has to be able to control all irrelevant variables and systematically manipulate the factors of interest. This, in turn, requires that the linguist knows details of the grammar or the logical structure of the language which are not accessible to naïve speakers. It is this knowledge, which is often drawn from existing theories, that allows the linguist to have intuitions about 'linguistic forms'. This leads us to Hintikka's key observation: competence theories do not have built-in devices for deriving abstract grammatical or semantic forms from particular linguistic samples; besides, 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 interdependencies 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. (52, p. 135)

If intuitions are to serve as a reliable starting point in linguistic inquiry, then 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 theories. The important 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 given by Hintikka to the problem of intuition points in the direction of *infusing linguistic practice with psychological experimentation*. The linguist would again 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 (see also Bunge (8, pp. 158-163) for a similar position).[17] Hintikka offers an intriguing example in which the imagined goal is to define the meaning of expressions in Montague grammar on the basis of 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 in the theory,

> 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. (52, p. 146)[18]

[17]Bunge (8, p. 168) pinpoints a few methodological choices within generative linguistics which seem to diminish its relevance in empirical science, such as the "conduct of linguistic inquiry in total independence from neuroscience, social science, and even scientific psychology" and "a heavy reliance on intuition". We also consider these as obstacles to understanding language, but we disagree with the severe judgment that Bunge draws from these observations, namely that modern linguistics is (or has been) pseudo-scientific. Compliance with standards of scientificity depends on several factors, such as for instance mathematical modeling, which in generative linguistics and formal semantics is highly developed and is contiguous with such established fields as mathematical logic and computer science.

[18]Although we consider Hintikka's an informative example of linguistic theorizing based on covariation patterns of contextual factors and intuitions, we must also add that there are serious problems with the notion of meaning (that is, Frege's *Sinn*) in Montague semantics. To mention one, since the theory allows for infinitely many possible worlds, it becomes problematic whether we can even approximate the meaning of an expression using Hintikka's method.

It is a fair point in favor of introspective judgments in a broad sense, however, to add that Hintikka considers thought experiments on a par with genuine experimentation on subjects (52, p. 146). Instead of eliciting overt responses from the subjects in a number of conditions, 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 systematic aspects of language and thus contribute to theories of competence.

Hintikka's argument can be made more explicit with reference to a number of studies investigating the role of the most important of his 'suitable ingredients' – context. Linguistic and psycholinguistic research demonstrated that the context in which a sentence occurs can affect judgments of acceptability. Bolinger (4) reported that sentences, which subjects judge as semantically implausible when presented in isolation, are regarded as acceptable when embedded in context. Consider the following examples:

(19) a. It wasn't dark enough to see.

b. I'm the soup.

These sentences are typically judged as semantically deviant, although for different reasons: (19a) since one normally needs light in order to see and (19b) because the predicate 'being a soup' cannot be applied to a human being. Now consider the same sentences embedded in a suitable discourse context, with (20b) being spoken at a cashier's counter in a restaurant:

(20) 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 (19a-b) in the appropriate context are perfectly acceptable. Since context has such marked effects on intuitions, linguistic theory, if it has to rely on introspective judgments, should explicitly take into account this fact. The same observation applies to experimental research on language processing, as we shall see below and in section 3.5.

3.2.2. *Intuitions in psycholinguistics.* The appeal to intuitive judgments was never an explicit methodological choice in psycholinguistics and the cognitive neuroscience of language. In fact, the method of introspection was discarded in scientific psychology after its failure in the 19th century. Nonetheless, possibly because in linguistics the method was regarded as sound, it was also adopted in language processing research as a means of establishing the relevant differences between sentence types to be used in actual experiments. The typical procedure in setting up a language processing study is to start with a small number of sample sentences differing with respect to some salient linguistic aspect, the evaluation of which is initially left to the intuitions of the experimenter. For instance, to consider one of the first ERP studies on syntactic constraint violations by Osterhout & Holcomb (80), the starting point would be a pair – or a relatively small set of pairs – of sentences containing either an intransitive (21a) or a transitive (21b) verb, using a direct object construction:

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

Up to this stage the methodology is by and large the same as that of the linguist. However, while the latter would then proceed with, say, formalizing the requirements of intransitive and transitive verbs with respect to direct objects, the psycholinguist, in order to ensure sufficient statistical power to test her processing hypotheses in a dedicated experiment, would have to construct a sufficiently large set of sentences with the same structure and properties of (21a-b). In the last step, the items would be presented to subjects while the dependent variables of interest are measured, which in the study of Osterhout & Holcomb were ERPs and grammaticality ratings. Now grammatical sentences like (21a) were judged to be acceptable in 95% of the cases and supposedly ungrammatical items like (21b) in 9% of the cases. One may argue, as a scholarly exercise toward an explanation of the 9% figure, that (21b) 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:

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

We have already encountered this phenomenon discussing Bolinger's examples above. In a context such as (22), Osterhout & Holcomb's ungrammatical sentence becomes perfectly admissible. A similar context of utterance might have been imagined by the subjects who judged the sentence acceptable. Needless to say, our observation does not affect the results obtained by Osterhout & Holcomb. What it shows, however, is that subjects' intuitions and judgments may differ to a variable extent from those of the experimenter. This is similar to the situation of section 2.3, where it was shown that subjects may interpret the reasoning task and the verbal material in a way that considerably differs from that of the experimenter.

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 roughly coincides with the dilemma as we described it above, between Chomsky's support of theoretical mentalism on the one hand and the choice of maintaining traditional linguistic methodology as based on intuitions and the competence-performance distinction on the other (55, p. 29). Accepting theoretical mentalism requires a revision or a rejection of the distinction between competence and performance. Jackendoff has addressed this issue, trying to find a more accommodating formulation which would allow a natural interplay of linguistics and the empirical disciplines investigating language processing. He writes:

> 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. (55, p. 33)

Jackendoff proposes to adopt a 'soft' competence-performance distinction, adding a third component to the framework (55, p. 34). In the new scheme, the theory of competence is the characterization of phonological, syntactic and semantic structures assembled and stored in the mind-brain of speakers during language acquisition. The theory of performance is the description of the occurrence of these structures in 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 on Chomsky's original distinction, it raises a number of problems. We start with how Jackendoff conceives the interactions between competence and performance.

In his recent book *Foundations of Language*, Jackendoff asks to what extent the logic of competence dictates the logic of processing (55, p. 200). Syntactocentric architectures, from the theory of *Aspects* (17) to Government-Binding Theory (18) and the Minimalist Program (19), postulate that recursion in narrow syntax is the only source of combinatoriality in the language system (51). Phonology and semantics are regarded as interpretive (as opposed to combinatory) systems. More specifically, phonology and semantics interpret a form which has been computed by the syntactic core in terms of non-linguistic structures, namely sound and conceptual representations. Phonology and semantics do not add combinatory structure to the syntactic representation. Instead, a semantic interpretation of a given sentence is derived by simply attaching lexical meanings to (some, but not all) the nodes of a syntactic tree; a phonological interpretation results from pairing a syntactic form with a sound representation. The logical directionality of the grammar is such that the syntactic form 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. For 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 comprehension and production is seen as a strength of the theory by generative grammarians. However, Jackendoff considers it a weakness, as it creates an undesirable and unnecessary gap between competence and performance. To fill this gap, he proposes an 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 key role in performance, acting as constraints applied in a parallel and interactive manner during comprehension and production, in a different temporal order in the two cases (55, 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. An example might help clarifying this point (55, p. 202): [revise based on Jackendoff 2007]

(23) a. It's only a PARent, not ∗ a TEACHer.

b. It's only apPARent, not ∗ REAL.

The two utterances up to the asterisk mark are acoustically indistinguishable, that is, word boundaries cannot be immediately established at the phonological level, but only when the subsequent segment is processed. It is the semantics of 'teacher' and 'real' that determines the meaning of the preceding clause, disambiguating the acoustic signal. Notice however that the semantic component cannot by itself reject the alternative phonological structure, because it does not contain rules that deal with the phonological input (55, p. 202). Thus, there must be interface rules allowing for semantic structures to serve as *additional* input for the phonological system. Jackendoff's specific proposal consists in treating lexical items themselves as interface rules between the conceptual-semantic and phonological systems. The lexicon as a whole is an interface component which, although it contains rules which are different from generative and transformational ones, is still part of competence (55, pp. 131, 425). Jackendoff takes (23a-b) as showing that the processor's 'opportunism'[19] is not chaotic, but is constrained by the structures available at a given processing stage and by the interactions between these structures via competence rules (55, p. 202).

Jackendoff provides an architecture in which the three competence components interact in a way that is consistent with the speed and accuracy of language processing. However, in order to solve the dilemma described above, establishing that competence determines the state-space available to language users during performance (55, p. 56) may not be enough. Rather, we want to determine whether there is genuine interplay between competence and performance, that is, turning Jackendoff's hypothesis upside down, whether the logic of processing dictates, to some extent, the logic of competence.

As we saw above, in his early writings Chomsky claimed that theories of competence have nothing to learn from processing or performance data (17). More recently minimalists have granted that syntax is adapted to the requirements holding at the interface with other cognitive modules, such as the sensory-motor and conceptual systems. Even so, they deny

[19] This is an informal term for the processing principle of immediacy, about which see (69).

what functionalists on the contrary accept, namely that syntax is well-designed for use or adapted to performance (20; 51; 31). Evidence against adaptation to performance would be provided, according to minimalists, by memory limitations, unstable constructions such as garden-path and center embedding sentences, and so forth. Here two remarks are in order. The first is that such phenomena do not constitute evidence against adaptation *tout court*, but rather (if anything like that is possible) against 'perfect' adaptation. Minimalists seem to commit what some optimality theorists have called a 'fallacy of perfection' (70), that is confusing optimal outcomes, which are the result of equilibria between different variables, with best possible outcomes for one or more of the factors involved, for instance absence of unstable or ambiguous constructions (see (86) for a discussion). The second remark is that, even if we assume that competence is neither perfectly nor optimally adapted to use, it is still possible that performance requirements shaped competence rules. Thus, the problem is not whether language is an adaptation: that some traits of competence reflect the outcomes of adaptive evolutionary processes acting on actual brain systems, including adaptation to communication needs, is a widely accepted view (51; 86; 31). The problem is rather that of developing a methodological framework in which it is possible to determine which aspects of competence can be explained adaptively and which cannot. This is one of the problems that one needs to raise in order to solve Chomsky's methodological dilemma.

The reason why generative linguistics does not seem capable of addressing this issue 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 has noted (55, p. 30), in Chomsky's original proposal a large and heterogeneous class of phenomena was collapsed into performance: errors, shifts of attention, memory limitations, processing mechanisms, and so on. Only a very superficial assessment of the factors involved in language use could justify the claim that a single, relatively compact theory of performance could account for all those phenomena. It is more reasonable to maintain that different theories, developed using different analytic approaches to the language system, are necessary to understand how competence interacts with memory and attention, how errors of different type and origin are produced (for also language disorders give rise to performance failures), and so forth. While we roughly agree with Jackendoff on the characterization of competence and neural implementation, we believe that a more appropriate intermediate level should be chosen. We present and motivate our choice below.

3.4. **Marr's three-level scheme as applied to language.** Spivey & Gonzalez-Marquez (98) have pointed out that Jackendoff's view of the competence-performance distinction as a soft methodological separation, plus the theory of the neural instantiation, resembles Marr's tripartite scheme for the analysis of cognitive systems (67, pp. 24-27). For our purposes, this is a rather perceptive remark, since it suggests that, while Jackendoff's refinement of Chomsky's distinction does constitute a progress, the trajectory is still incomplete. Here we propose to replace the competence-performance distinction with Marr's scheme, at least to face the specific problems we have been discussing in the last few sections. One important consequence of this move is that we will replace Jackendoff's performance theory with an intermediate level of analysis at which the algorithms and memory mechanisms supporting the relevant computations are described. This might seem a rather abrupt move, especially because it restricts the scope of performance to algorithms and thereby disregards a large number of phenomena which, some may argue, cannot be treated in computational terms. This is admittedly true. Nonetheless, one advantage is that, for those phenomena that allow for a computational treatment, such as language processing, we have a framework in which competence theories, properly construed, can be investigated as part of actual information processing systems. As we shall demonstrate later on, this can be beneficial for some areas of psycholinguistics and the cognitive neuroscience of language. We will now apply Marr's scheme to language to lay down the basis of our proposal.

The first level, which Marr & Poggio (67; 68) call the level of the 'computational theory',[20] contains a characterization of the data structures which the language system constructs and on which it operates during acquisition, comprehension and production. Computational analyses include also a description of the assembly rules and basic constituents stored in

[20]This appears a rather infelicitous terminological choice, as 'computational theories' in Marr's sense have to do more with a description of the *goals* of the computations rather than with the computations themselves. The latter are described at the algorithmic level of analysis or, if we accept the notion of 'computation' used in neuroscience, at the level of the physical implementation, where low-level mechanistic models are provided. However, we find it hard to come up with a better term, and therefore we will stick to Marr's own choice. Unless specified otherwise, whenever we use the term 'computational', we use it in Marr's sense.

memory. For instance, a computational theory for syntax will contain a characterization of both phrasal constituents and attachment rules (see below 3.5.2). A computational theory for semantics will contain a formal characterization of the classes of structures in which the relevant expressions, translated into some suitable formal language, can be interpreted, as well as a definition of valid inference (102). Computational analyses in Marr's sense and competence theories in Chomsky's loosely coincide, with two important differences. First, computational theories do not make any claim as to the particular form in which linguistic principles are stored in the mind-brain of speakers, while linguistic competence is typically taken to be a form of tacit knowledge. Second, the notion of 'information processing goal' is accommodated within the computational theory and thereby within linguistics. The idea is to take the computation of a particular linguistic structure, for instance a parse tree or a model verifying a given sentence, as a *goal* – possibly involving subgoals – which has to be attained by the system during comprehension and production (50; 2). This brings us back to the conjecture that language processing has an intimate connection with action planning. As we have seen in section 2.2, a sentence S given as input can be seen as an instruction for the system to attain a particular goal: constructing a discourse model in which S is true. Notice that computational analyses specify only what these goals and subgoals are, not the processing steps required to update the discourse model accordingly. This is dealt with by the theory of algorithms (more on that below). Computational theories are inspired by and testable on behavioral data (68, p. 7), although intuitions are also admitted within Hintikka's conditions. Analyses at the first level are designed to contribute to cognitively plausible theories of language. An important requirement is therefore that the goals they describe are effectively computable.

The second level contains the theory of algorithms and abstract memory mechanisms. In his 1982 book *Vision*, Marr adopts a single level at which only algorithms are described. In a 1976 paper, however, Marr & Poggio (68) split the intermediate level of analysis into two, with the result that memory mechanisms are handled independently from algorithms and mediate their implementation into the neurobiological system. Here we follow Marr's 1982 proposal, assuming that algorithms and memory mechanisms can be described at the same level of analysis. The main reason for this choice is that the same techniques can be used to test algorithms and memory mechanisms. Behavioral measures such as reading and reaction times, eye movements, and neurophysiological ERP or ERF data seem adequate for this purpose since they typically afford reliable inferences concerning the complexity and the time course of linguistic processes, which in turn may be used to determine whether computation is serial or parallel, immediate or delayed, and so forth. Cognitive processes can be modeled within the Church-Turing framework of computability. The computation of linguistic structures in the brain can be represented using a set of finite-state machines, combined with a flexible memory organization (113; 84). Therefore, analyses at Marr's level 2 can be formulated in terms of dynamical systems theory (83; 84).

The third level contains the theory of the neural implementation. Here analyses of brain systems carried out at different levels of granularity are required. Coarser-grained analyses amount to descriptions of the spatiotemporal dynamics of language processes as revealed by techniques such as EEG, MEG and fMRI, integrated by patient data (7). These accounts may feature statistical maps displaying the functional and anatomical connectivity between the different brain systems involved in language processing, as well as between language and other cognitive domains, such as perception, action, memory and executive control (84; 45). Finer-grained analyses amount to descriptions of the properties, connections and activation patterns of the neuronal populations forming the structures under exam. Neural networks, in particular connectionist models, may provide a first approximation toward biologically plausible accounts of induced and evoked neural dynamics (see (101; 102; 2) for applications). The motivation for an independent level of neural implementation is not to characterize the empirical domain of computational or algorithmic theories. In Marr's scheme, the relation between theory and observation is not articulated along the 'vertical axis' of levels of analysis. The level of the biological implementation contains mathematical models and hypotheses to be tested on various data types, just as computational theories are evaluated on the basis of behavioral data, and algorithms are tested on reaction and reading times, ERPs, and so on. The result is a unifying framework which allows to specify along the 'vertical axis' the relations between linguistic theories, processing algorithms and models of neural computation, and along the 'horizontal axis', for each of these levels of analysis, the appropriate empirical domain of verification.

The issue which should now be addressed is whether our appeal to Marr's scheme solves the problems associated with the competence-performance distinction. One improvement is that the different phenomena that were collapsed into the notion of performance can now be understood in their distinctive features. For instance, working memory as involved in a given process can be examined at the intermediate level. Algorithmic analyses may lead to a description of the minimal memory architecture and memory resources required by the system, which constitutes a first step toward an explanation in terms of neural processes. Conversely, memory organization constrains the type of algorithms that can be computed by the system and thereby the type of data structures that the computational theory can posit. An example of such bottom-up adjustment is Yngve's explanation of the dominance of right-branching over left-branching and center-embedding structures (119). The notion of 'minimal model' (see (114) and 4.2.2 below) in the Event Calculus treatment of tense and aspect is an example of accommodating a resource-sensitive data structure within the competence theory. Taking into account memory limitations and other constraints should improve the overall theory by achieving a sort of 'reflective equilibrium' between analyses at different levels. This gives a tentative answer to the question to what extent the logic of processing dictates the logic of competence, as it points to one well-defined bottom-up constraint: effective computability.[21]

The existence of bottom-up constraints on theories of competence, however, does not undermine the independence of top-level computational analyses. Marr has emphasized 'the importance of computational theory' (67, pp. 27-29): without an independent *logical* description of the problem to be solved, the design of the algorithms can only proceed in a bottom-up fashion, inspired by an intuitive notion of what the system should compute, but essentially adjusting the parameters of the 'computational model'[22] until it conforms to the observed data. Two problems. First, internally consistent models which approximate the behavior of the system are not necessarily correct. The realizability in computer programs often guarantees a theory's local consistency, but by no means guarantees its correctness.

[21]For a discussion of computability and examples of the use of bottom-up constraints, see (74; 83; 84; 113).

[22]Here in the sense of cognitive science: a neural network model of the process of interest, rendered as a computer program simulating the behavior of the system.

Superficially, successful computational models can be the undoing of the science they purport to aid. The ability of a model to deal with a large but manageable number of isolated facts has little to do with having a correct theory. In fact, factual success can unnaturally prolong the life of an incorrect theory. As the unknown wag says, "if Ptolemy had had a computer, we would still believe that we are the center of the universe" (107). Second, this approach might provide models which efficiently compute some very specific function (e.g. the degree of relatedness between the meaning of an expression and its semantic context) but fail to account for how the computation of that specific function contributes to solving the overall information processing problem (e.g. computing a model of the discourse in which the expression occurs) and how the same function can contribute to the solution of other problems (29). In order to integrate the output of each particular algorithm into the solution of the overall computational problem, it is necessary to specify 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. Marr's functionalism could infuse cognitive science with a deeper and more articulated notion of computational modeling, starting from a relatively 'simple and clean' logico-mathematical account of the overall computation, then implemented into a set of algorithms and neural mechanisms effectively carrying out the computation, and back again. Most importantly, the predictive and explanatory powers of the theory might increase if an abstract computational theory is specified. Below we provide two examples suggesting that this might indeed be the case.

3.5. **Two experiments on semantic processing.** In this section we review two studies on semantic processing, with two main purposes: the first is to provide support for the claims about intuitions and levels of analysis made above; the second is to introduce the models of syntactic and semantic unification discussed from a wider theoretical perspective in section 4. The ERP studies by Nieuwland & van Berkum (78) and Münte et al. (76) are particularly interesting for our purposes as they provide accounts of experimental results which, upon a closer look, have at least equally strong competitors based on more explicit computational models of language processing. Proposing an alternative explanation without conducting an experiment on the basis of which the incorrect hypothesis (if any) can be ruled out does not disqualify the original account. Nonetheless, reasoning on alternative explanations can be a very instructive exercise, and as such this section is intended to be read. The aims of our exercise are (a) to evaluate the status of the original accounts, in particular their predictive and explanatory powers, by contrasting them with alternative analyses and (b) to illustrate the benefits of more explicit computational analyses on experimental research. Let us first introduce the technique with which these two experiments were carried out.

3.5.1. *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 (64). The resulting electroencephalography (EEG) is thought to reflect the activity of a large number of cortical sources. Only a small part of the electrical brain 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 methods for extracting event-related potentials (ERPs) from the EEG. Averaging over a relatively large number of trials of the same condition is the most widely used approach. The assumption is that noise, here 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 tend to reduce noise and reveal the ERP signal.[23] ERPs seem particularly suited for the study of information processing as they provide an excellent window on the temporal dynamics of neural activity (26). 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 (64).

Event-related potentials have proved useful to address a number of issues concerning the relative time course and complexity of phonological, syntactic and semantic processes. Kutas & Hillyard (59) conducted the first ERP experiment in which linguistic factors were successfully manipulated, in the specific case the semantic plausibility of a word given the preceding sentence context:

(24) a. The officer shot the man with a gun.

b. The officer shot the man with a *moon*.

[23] For details on obtaining ERPs, see Luck's book (64, Ch. 1, 4).

Compared to the control 'gun', the anomalous noun 'moon' elicited a larger negative shift starting around 250 ms after word onset, peaking at 400 ms and lasting until approximately 550 ms. The component, which was called N400 because of its polarity and peak latency, is known not to be affected by unexpected events such as variations in the physical properties of the stimuli, for instance the size of the letters presented on the subject's screen. Larger N400s are evoked also by semantically plausible words which are nevertheless judged by subjects as less preferred in a given sentence context (60; 46), like 'pocket' below (fig. 2):

(25) a. Jenny put the sweet in her mouth after the lesson.

b. Jenny put the sweet in her *pocket* after the lesson.

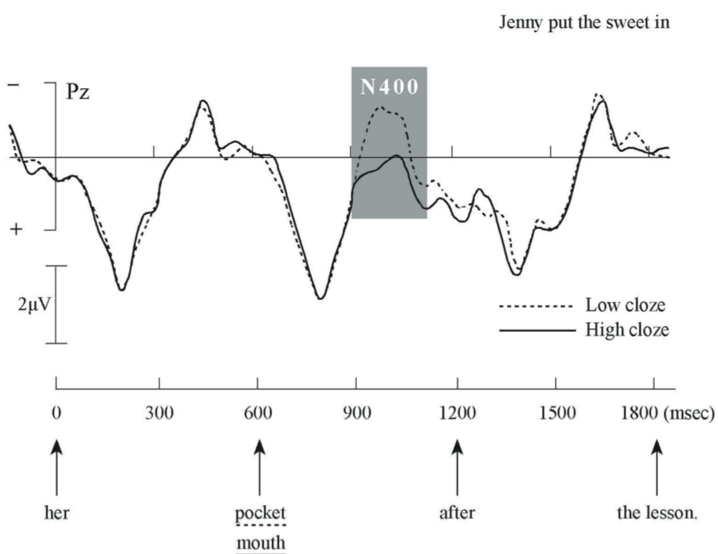


FIGURE 2. N400

The amplitude of the N400 is also modulated by lexical items which provide information conflicting with the preceding discourse context (111; 112) or with world knowledge (49). In summary, although every content word evokes an N400, the component is affected by the degree of semantic fit of the lexical item with the preceding context and the knowledge base made salient for its integration.[24] Semantics-related negative shifts different from the N400

[24]Recent studies suggest that larger N400s are also elicited by co-speech gestures which do not match with the semantic content of the accompanying sentence. This indicates that the amplitude of the N400 indexes also difficulty in cross-modal semantic integration. See the study of Özyürek et al. (81) for experimental data and discussion.

are elicited by constructions considered taxing on working memory such as referentially ambiguous NPs (110; 109) and temporal connectives (76), to which we will return later.

An ERP component associated with syntactic processing was reported independently by Osterhout & Holcomb (80) and Hagoort et al. (48) in sentences violating grammatical constraints such as:

(26) 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 verb 'is', the incorrect verb form 'are' elicited a larger positive shift starting around 500 ms following the onset of the verb and lasting for about 500 ms. The component was called P600 (80) after its polarity and peak latency and SPS (Syntactic Positive Shift) on the basis of its functional properties (48). Larger P600/SPS components are also elicited by grammatically anomalous meaningless sentences (46) (fig. 3):

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

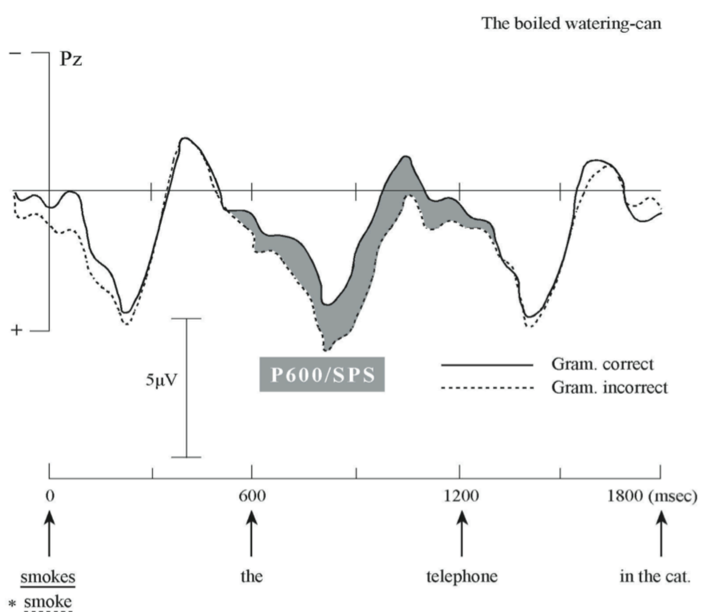


FIGURE 3. P600/SPS

A larger P600/SPS is also evoked in temporarily syntactically ambiguous sentences by the word which introduces the only parsing option available, like ‘noticed’ in (28b) compared to the unambiguous sentence (28a) (47):

(28) 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 (28a) ‘the cowboy’ can only be the subject of the subsequent VP, in (28b) it could also be a direct object of the preceding VP until ‘noticed’ is encountered. Disambiguating a syntactic representation thus results in a larger P600/SPS. Efforts toward a computational framework accounting for syntax-related ERP effects have been made by Hagoort (42) using a lexicalist parser elaborated by Vosse & Kempen (117), which we will see applied below.

3.5.2. *Semantic illusions.* Nieuwland & van Berkum (78) were interested in the ERP effects of coherence breaks in discourse. They presented subjects with short stories of four sentences (29a) 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, the woman continued the conversation either with the man as in (29b) or with the inanimate object as in (29c). The noun referring to the man in the control sentence was replaced by the one referring to the object in the critical item:

(29) 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 (29c) marks a strong coherence break in the story. Accordingly, an N400 should be expected at the critical word. However, the N400 elicited by ‘suitcase’ was not larger than the N400 evoked by ‘tourist’. Instead, a larger P600/SPS was observed at the critical word. Presenting (29b) and (29c) 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 recorded data to the *absence* of the process of interest. They take a null effect in the N400 time window as evidence that the replacement of 'tourist' with 'suitcase' did not result in differential processing. Their argument is that, since even very subtle lexico-semantic differences affect the amplitude of the N400, the absence of a modulation of the component indicates that the replacement went undetected. Regrettably, the authors do not provide behavioral evidence that subjects failed to notice the coherence break, comparable to the well-known data on the 'Moses illusion' (27) to which Nieuwland & van Berkum also refer. Without behavioral data, one might also speculate that the anomaly *was* detected and that this did result 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 more explicit theory 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 substitution of referents while the latter effect cannot be elicited in such circumstances. Without this sort of analysis, 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. Moreover, without a computational analysis, it remains unclear whether and how exactly a short story in which the discourse referents are mentioned several times can induce 'semantic illusions' as, according to the authors, would be suggested by the fact that a larger N400 is elicited by the noun 'suitcase' only when the sentences are presented in isolation.

Nieuwland & van Berkum consider two possible accounts of the observed P600/SPS:

> One potentially viable way to resolve the apparent conflict in obtaining a P600 effect to a – not initially detected – semantic anomaly is to assume 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). Alternatively, if the P600 more generally indexes a monitoring process that checks upon the veridicality of one's sentences perception, as has been recently proposed, the conflict would also be resolved. (78, p. 699)

There appears to be no principled way to derive these two hypotheses, either from linguistic theory or from current knowledge of the P600/SPS. If, following the first suggestion, the P600/SPS reflects a syntactic response to a semantic comprehension problem, 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 reasonable conclusion at this point 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 interpreted as a signature not of syntactic processing but of a 'monitoring process' or, as Nieuwland & van Berkum also suggest (78, p. 699), as an index of a "delayed and more careful interpretive process that subsequently overrides [the initial] incomplete, illusory interpretation", then one should be prepared to accept the more parsimonious interpretation of the P600/SPS (equally distant from standard syntax-based accounts of the positive shift and presumably equally implausible) as simply reflecting the detection of a coherence break. But even if we accept the explanation of Nieuwland & van Berkum, it remains doubtful whether the P600/SPS could be a 'delayed and more careful' response to the anomaly. There is no criterion as to what counts as a delayed response, but the few hundreds of milliseconds separating the N400 (which, according to the authors, would be the appropriate response, thus not a delayed one) from the P600/SPS may not be enough. In addition, if the response was only delayed, an ERP effect in which the polarity is reversed with respect to the N400 should not be expected. Indeed, the observed response was *qualitatively* distinct, not just delayed. Below we try to outline a computational analysis of the observed P600/SPS, based on two ingredients.

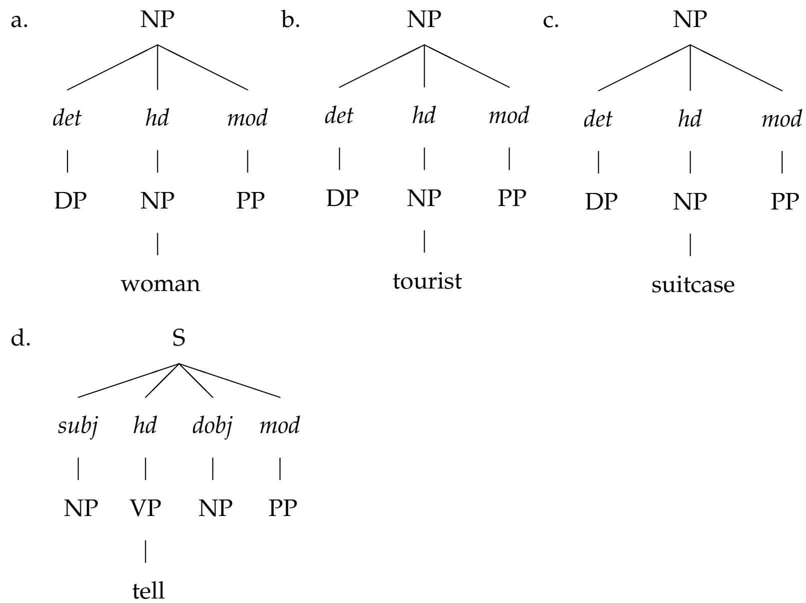
The first ingredient is borrowed from semantic theory. Sentences like (29c) can be seen as instances of *metonymy* or *coercion*, where the noun 'suitcase', normally referring to the inanimate object belonging to the man, is used to refer to its owner instead. Nieuwland & van Berkum dismiss these constructions because they "derive much of their attractiveness as stylistic or humoristic devices from the very fact that they are less expected" (78, p. 699). This, however, appears to understate the range of the appropriate contexts of use of such expressions. As a matter of fact, coercion is used in many more cases than the humorous or stylistically marked discourses alluded to by Nieuwland & van Berkum. The following examples are taken from Jackendoff (55, p. 388):

(30) 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.

In the first example, which may or may not sound humorous, possibly depending on other contextual variables, coercion might be used to deal effectively with a situation, such as a busy restaurant, in which utterances should be maximally brief and informative. Sentences like (30b), used in the context of a bookstore or a library as an indication of where to find the desired books, might be preferred to lengthier descriptions such as 'The works of Plato are on the top shelf next to those of Lord Russell'. Anyone with some experience of libraries or bookstores would agree that such shorthands are in fact relatively frequent. Analogous considerations hold for (30c), which exploits the proper name and the possessive pronoun to construct an efficient shorthand, without being either humorous or stylistically marked. What enables coercion in all these cases is *context*: if the information available to the hearer makes clear which entities the speaker is referring to, such constructions are permissible and, when pragmatically more efficient, also preferred. Coercion in context is thus the first ingredient of our account.

The second ingredient is taken from a lexicalist parser developed by Vosse & Kempen (117) and adapted by Hagoort (42) as a computational analysis 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. Consider the example (29a-c) above. When 'woman', 'tourist', 'suitcase' and 'told' are processed, the syntactic frames (31a-d) become part of the unification space. The top layer of each 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 linked. The nodes in the second layer depend on the phrasal constituent occupying the root node. For instance, an S node will have four functional nodes specifying its permissible syntactic environment, that is subject, head, direct object 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 syntactic frame can be attached.

(31)



Syntactic unification consists first in checking agreement and other features of words such as 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 (see (42; 117)), but it may be useful to mention that unification is based on competition between alternative attachments and lateral inhibition to suppress discarded options. Hagoort (42) proposed that the amplitude of the P600/SPS is related to the *time* required to establish attachments of sufficient strength. This temporal variable is affected by syntactic ambiguity (amount of competition), syntactic complexity, semantic influences and failed binding attempts as in syntactic violations.

The model just presented suggests the following computational analysis. The NP frames (31a-c) are introduced in the unification space when (29a) is processed. Sentences (29b) and (29c) introduce the additional S frame (31d) for the verb 'told'.[25] In (29b), the NP root node of 'woman' is bound to the identical foot node of 'told' as the subject of the verb phrase.[26] When the noun 'tourist' is encountered, successful animacy check will allow its root node to be bound with the remaining NP foot node of the verb. Unification is not as smooth in

[25] To avoid complicating matters, we leave tense aside here.

[26] For simplicity, we leave out binding operations for determiners.

the critical item (29c). Also in that case, the root node of 'woman' is bound to the subject NP foot node of 'told'. When 'suitcase' is encountered, animacy restrictions will prevent its root node from binding with the remaining NP node of 'told'. At this point (and here comes our hypothesis) discourse semantics is recruited to solve the impasse. 'Suitcase' will be treated as a coercing NP, referring not to the inanimate object, but to the man carrying it. Coercion is permissible because the conceptual relation between the tourist and the suitcase has been introduced in the discourse model by (29a), and therefore it is available to later processing stages. As soon as the new denotation for 'suitcase' has been computed, syntax can take up again the binding work from where it had left it, and unify the foot and root NP nodes of 'told' and 'suitcase'. Interrupting syntactic unification, establishing the new denotation for the coercing NP and resuming the binding process takes time. On Hagoort's (42) account of the P600/SPS, the positive shift reported by Nieuwland & van Berkum would reflect the longer time required to bind the verb 'told' to the coercing noun 'suitcase'.

Computing a new denotation for 'suitcase' is possible only if context licenses the proper *semantic* relation between discourse items, that is only if the information that the tourist was carrying the suitcase is available in the discourse model. The consequence is that, when the required semantic relations are not provided by the context, coercion is not permissible. So if (29c) was processed without the information that the man carried a suitcase, the parser would have no alternative option than attaching the root NP of 'suitcase' with the foot NP direct object node of the verb. In that case, syntactic unification would proceed seamlessly, although a violation of semantic restrictions (animacy) would be obtained.[27] This explains why the P600/SPS was not elicited by (29c) sentences in isolation and why an N400 was obtained instead.

The analysis sketched above is not intended to be taken as a detailed, fully-constrained computational model – it clearly is not. Still, it may help us to evaluate the explanatory and predictive powers of the original hypotheses of Nieuwland & van Berkum. First, it shows that the P600/SPS can be accounted for in a more parsimonious way, without postulating a 'semantic illusion' effect which, given the absence of behavioral data for the relevant cases,

[27]Notice that when (29a) is presented in isolation there can be no coherence break, as there is no context setting up coherence requirements. For that reason, the N400 observed by Nieuwland & van Berkum was more likely due to an animacy violation rather than to a coherence break.

remains as yet undemonstrated. Second, it suggests that an explanation of the P600/SPS in terms of syntactic binding operations is possible, avoiding revisionary interpretations in terms of 'monitoring' and 'delayed' responses. Third, it shows that in processing the critical item (29c) semantic computation is not disrupted and that, therefore, there is no reason to expect an N400. With the exception of (29c) presented in isolation, animacy restrictions are not violated, but are actively used to reject the initial attachment of 'suitcase' as the direct object of the verb. Discourse semantics is not 'violated' either. Instead, contextual cues are invoked to select the appropriate reference for 'suitcase'. Fourth, testable predictions can be derived from our analysis: in metonymy or coercion cases like (30a-c) above, a P600/SPS is expected when context licenses coercion, an N400 when no alternative unification option is left to the parser. In conclusion, even a relatively informal computational analysis may give more definite insights into what an experiment actually shows and what can be expected from further work. We provide another example below.

3.5.3. *Temporal connectives.* Münte et al. (76) recorded ERPs while subjects read sentences differing only in the initial temporal connective, for instance:

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

An additional test assessing participants' working memory span was performed. Subjects read aloud increasingly larger 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 then used to group subjects (high, medium and low working memory span). 'Before' sentences elicited a larger sustained negativity, maximal over left anterior sites. At the left frontal electrode, ERP responses to 'before' and 'after' diverged as early as 300 ms after sentence onset. The effect lasted throughout the sentence and was larger during the second clause. The difference of left anterior negativity between 'before' and 'after' sentences was positively correlated with working memory spans: 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 as follows. As reported above, ERP waveforms 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 memory: in sentence-initial position, 'after' indicates that the events are presented in their actual order; 'before' indicates that the order is reversed. As regards the larger sustained negativity elicited by 'before' sentences, it is hypothesized that these require additional computations as a consequence of presenting the events in reverse temporal order.

An important assumption, upon which the account of Münte et al. ultimately relies, is that the order in which the events are mentioned in discourse is the only difference between 'after' and 'before' sentences. This follows from the claim (76, p. 71) that what is mapped onto the meanings of 'after' and 'before' are relations of temporal ordering of events, that is temporal precedence and temporal consequence. A review of the linguistic literature on the subject, however, suggests that there is more to the semantics of temporal connectives than event order. There is indeed an asymmetry between the inferences licensed by 'after' and 'before': while 'after' always entails that the subordinate clause is true, 'before' allows for the subordinate clause to be either true or false, depending on the information provided by the main clause (1; 23). Example (32b) suggest that the content of a 'before' clause, if the earlier event normally prevents the later one from occurring, will be interpreted as false or, more precisely, as non-veridical. In the situation described by (32b), 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 sentence, but this is quite different from a plain temporal ordering of the events.

Our analysis is based on the mathematical theory of tense and aspect proposed by van Lambalgen & Hamm (114), which has been briefly described and applied in sections 2.1 and 2.2. In that framework, tense, aspect and temporal adjuncts are represented by integrity constraints. Recall that these are instructions in the form of obligations and prohibitions to update the current model such that a representation making discourse true is incrementally computed. 'After' sentences must satisfy three integrity constraints. Two are introduced by the subordinate clause and require an update of the unification space such that the relevant activity (the submission process) holds in the past of the moment of speech and is finished before t'', where the event occurring at t'' is to be specified by the subsequent main clause:

(33) a. $?HoldsAt(submit,t) \wedge t < now$ `succeeds`

b. $?Happens(finish,t'), t' < t'' < now, Happens(e',t'')$ `succeeds`

The variable e' denotes an event type and has to be unified with the event described by the main clause. Here *finish* is the culminating event, and is followed by the goal state, that is a complete submission. The main clause introduces the second integrity constraint:

(34) $?Happens(change,t''), t'' < now$ `succeeds`

where *change* denotes the journal's policy change. The unification $e' = change$ leads to

(35) $?Happens(finish,t')$, $t' < t'' < now$, $Happens(change,t'')$ `succeeds`

This is the only binding operation required by 'after'. It also follows from (35) that 'after' sentences license veridical readings only.

'Before' sentences introduce three analogous conditions. The first two are added to the unification space by the subordinate clause; the second is derived from (33b) by replacing the obligation with a prohibition. In this way, we account for 'after' and 'before' as denoting converse temporal relations:

(36) a. $?HoldsAt(submit,t) \wedge t < now$ `succeeds`

b. $?Happens(finish,t'), t' \leq t'' < now, Happens(e',t'')$ `fails`

Also in this case e' is an event type to be unified with a token described by the main clause. The third condition introduced by the main clause is identical to its 'after' counterpart (34). The system will also in this case unify $e' = change$, reducing (36b) to the request

(37) $?Happens(finish,t'), t' \leq t'' < now, Happens(change,t'')$ `fails`

Notice that (37) can be satisfied in different models, since the integrity constraint fails if and only if *at least one* of its conjuncts fails. Suppose first that there is no information in semantic memory to the effect that a journal's policy change is a likely obstacle toward a submission. Closed world reasoning yields $\neg Terminates(change,submit,t)$. From this it can be concluded that the submission was completed, because the process took place as is required by (36a) and it can be assumed to be finite. Given that both events (the complete submission and the policy change) happened, the only solution left for the system is to add $t'' < t' < now$ to the unification space in order to ensure that (36b) fails. This corresponds to the veridical reading in which the submission follows the journal's policy change. Suppose however that

$Terminates(change, submit, t)$ is retrieved from semantic memory. The submission process is interrupted before it is completed. This implies that the canonical terminating event did not take place and that therefore the goal was not attained. The first conjunct of (36b) fails, yielding a non-veridical reading.

In order to consider the implications of our analysis at the algorithmic level we would need the full computational machinery of van Lambalgen & Hamm (114), combined with hypotheses on the time-course of linguistic processes, such as for instance immediacy (69). However, here we can only give a qualitative description of the processing steps involved in processing 'after' and 'before' sentences. Furthermore, we will sketch a processing account which can be reconciled with the results of Münte et al., warning the reader that alternative reconstructions are possible.

Processing 'after' and 'before' sentences amounts to incrementally computing discourse models which satisfy (35) and (37). Processing 'after' clauses consists in computing a model in which (33a) and the first two conjuncts of (33b) are satisfied – our version of immediacy implies that processing algorithms can satisfy incrementally the conjuncts of an integrity constraint of the form $?\varphi(x)_1, ..., \varphi(x)_n$ `succeeds`. Processing the main clause amounts to forcing an update satisfying (34) and unifying event variables, which automatically leads to a model satisfying (35). Processing 'before' clauses starts with the computation of a model in which (36a) can be satisfied. At that stage, however, none of the conjuncts of (36b) can be satisfied, for the integrity constraint is of the form $?\varphi(x)_1, ..., \varphi(x)_n$ `fails` and as such can lead to very different updates of the discourse model. The algorithms which should update the model according to the information provided by the subordinate clause must wait for the main clause to be interpreted; the main clause, and related causal knowledge retrieved from semantic memory, determines whether the first conjunct of (36b) can be satisfied and thus the second must fail (as in a veridical reading) or whether it is the first conjunct that fails (as in a non-veridical reading). Because the execution of these algorithms is delayed, the integrity constraint (36b) must be maintained in working memory until relatively late before it can be satisfied. This might explain why 'before' sentences are more taxing on the memory resources available for unification, especially during later processing stages when the interpretation of the two clauses has to be finalized in a coordinated manner.

The explanation just sketched shares with the account of Münte et al. the hypothesis that the order in which events are mentioned in 'before' sentences may affect neural processing. Yet our analysis appears to be more comprehensive, for it takes into account veridicality, a factor which might contribute to understanding the sustained negativity reported in (76). Although in this case the semantic analysis simply enriches the original proposal of Münte et al., our exercises in this section demonstrate the benefits of more explicit computational theories in Marr's sense for language processing experiments. Further, using semantics to understand ERP data constitutes an example of how a coordination of different theoretical and experimental endeavours in the cognitive science of language can be achieved. Below we provide additional motivation for a comprehensive approach to language and we chart some directions of future research. We start with a discussion of the 'binding problem' for language, which we extend to the relatively unexplored case of semantic unification.

4. THE ROAD AHEAD

4.1. **The binding problem for language.** The first formulation of the 'binding problem' is due to von der Malsburg (115), who considered the binding approach to brain function as a response to some difficulties encountered by classical neural networks. Von der Malsburg (116) refers to a well-known example by Rosenblatt (94) to illustrate the nature of the issue. Imagine a network for visual recognition constituted by four output neurons. Two neurons fire when a specific 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 rectangular frame). For instance, if there is a square at the top, the output will be [square, top], if there is a triangle at the bottom, the output will read [triangle, bottom]. However, if a triangle and a square are presented simultaneously, say, the triangle at the top and the square at the bottom, the output may be [triangle, square, top, bottom], which is also obtained when the triangle is at the bottom and the square at the top: the network responds adequately only when a single object is presented in the rectangular frame. This is an instance of 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. (115, p. 96)[28]

Examples of the binding problem are 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, depending on how these features are bound together. Since the availability of different representations essentially depends on the geometric properties of the figure, rather than on the constitution of the perceptual system as it would be the case, for example, for after images (67, pp. 25-26), bistability requires an explanation at Marr's computational level, where properties of stimuli are described and related to information processing goals. Without a characterization of the geometric properties of the figure and of the mappings between the figure and the two different entities (for instance, a duck and a rabbit) which it can stand for, there would be no basis upon which to claim that the two representations are mutually exclusive. This suggest that binding in the visual domain is a genuine information processing or computational problem in Marr's sense. But what about language?

Although binding mechanisms are likely to differ across cognitive domains (43), there exist analogous cases of structural ambiguity in language. Take the following sentences:

(36) a. The woman saw the man with the binoculars.

b. Respect remains.

Example (36a) has two alternative syntactic representations, one in which the phrase 'with the binoculars' is a PP attached to the NP 'the man' (the man that was seen by the woman had binoculars), and another in which it modifies the VP (the woman used binoculars to see the man). Also in this case, the features of the stimulus lead to two distinct representations, depending on which attachment option is eventually pursued. These items typically result in larger P600/SPS components, suggesting that syntactic binding is a genuine information

[28]Different solutions to Rosenblatt's problem are possible. See von der Malsburg (116) for a proposal in line with the binding hypothesis and to Riesenhuber & Poggio (92) for an alternative approach.

processing problem for the brain. Sentence (36b) has two possible interpretations: on one it is a request, where 'respect' is the verb and 'remains' is the object noun; on the other it is a statement, where 'respect' is the object noun and 'remains' is the verb. There may well be some superficial similarities between perceptual and linguistic bistability, such as the fact that in both cases we seem to 'flip' between the two incompatible representations. But there is also a deeper analogy between the two: structural ambiguity is defined at the topmost level of analysis in both cases, as Marr himself pointed out (67, pp. 25-26). Without such an independent characterization it would remain unanswered why such representations are mutually exclusive in the first place. To extend Marr's line of reasoning, we suggest that the binding problem for language is best formulated at the computational level, although attempted solutions are bound to require significant adjustments at all levels of analysis, including – perhaps most interestingly – the level of neural implementation (43; 44). While there exists a computational formulation and solution of the binding problem for syntax (42), which we applied in section 3.5.2, a parallel piece of work for semantics is still missing. Our next step is to motivate and outline an account of semantic unification.

4.2. **Toward a model of semantic unification.** The separation between formal semantics and the empirical study of language comprehension has persisted from the early days of Frege until fairly recently (50). For many different reasons, ranging from philosophical to technical, the coordination, to say the least, of semantic theory and the cognitive science of language seems as desirable as is difficult to achieve. There are obstacles on both sides of the fence, with semanticists proposing theories based on not necessarily computable functions, such as for instance the treatment of Frege's notion of sense in possible worlds semantics, and psycholinguists adopting obscure 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. However, it seems possible to list some of the main ingredients of a model of semantic binding, without giving a recipe for cooking them up. We focus on three such ingredients, namely a formal account of unification, a cognitively realistic notion of meaning, and non-monotonic inference.

4.2.1. *Integration, binding and unification.* Unification in the sense described in 2.2 and 3.5 differs from the standard notion of 'integration' in that it is formally defined as an operation on syntactic frames and event variables, and has a role in the computational machineries of lexicalist parsers (117) and Event Calculi (114). 'Binding' may involve unification, but in general it refers to syntactic and semantic unification in a more 'hostile environment', when structural ambiguity is present. For instance, unification of syntactic frames for (36a) is an instance of binding, whereas for (29c) it is not. In conclusion, unification should not be confused with integration and binding, although these notions are closely related.

Unifying event variables has on semantic representations the same effect that unifying root and foot nodes has on syntactic frames in the model of Vosse & Kempen (117). That is, it leads from a set of sparse semantic representations, each with its own set of variables and event types, to an overall discourse representation in which variables occurring in different integrity constraints refer to the same processes or events. We have stressed how important unification is in order to relate the events described by sentences like 'Max fell' and 'John pushed him', and to relate the temporal content of main and subordinate clauses in 'after' and 'before' sentences. Further, we have said that, in both cases, unification builds on causal or world knowledge and that inference may be required in order to bring such knowledge to bear on discourse interpretation. As we have noted in section 1 that there exist semantic and psycholinguistic theories that regard inference and the integration of knowledge from different sources as a step *which follows and builds upon semantic interpretation* (57). One upshot of our account is that, contrary to what these approaches maintain, unification (and hence interpretation) would be hardly possible without inference on world knowledge. The examples on Max and John, French tenses, and temporal connectives discussed above lend some plausibility to this view.

Our hypotheses on semantic unification may be used to account for modulations of the amplitude of the N400. Echoing the explanation of the P600/SPS adopted by Hagoort (42), we propose that the amplitude of the N400 depends on the amount of competition between alternative unifications of event or fluent variables, and on the time required to establish unification links of sufficient strength.[29] Consider the following examples:

[29] While the model of the P600/SPS of Vosse & Kempen (117) and Hagoort (42) is based on precise notions and measures of competition and unification strength, we have nothing comparable for the N400 as yet. However, logic programming as described in (114) can be systematically related to spreading activation networks (101; 102), which should make the task of providing such definitions and measures feasible. See (2) for progress in this direction.

(37) a. The man read the book about the Dutch revolution.

b. The man *finished* the book about the Dutch revolution.

Sentences like (37b) are usually referred to as instances of coercion, of a different type than the cases discussed in 3.5.2 though. Since (37b) is not explicit as regards to what activity was finished by the man (reading or writing), the reader has to reconstruct the appropriate event sense on the basis of her knowledge of the agent and the object involved. In the Event Calculus (114), the coercing VP *finish* can be represented as the integrity constraint:

(38) $?Happens(e,t), Terminates(e,f',t), t < now$ `succeeds`

where $e = finish$ and the variable f' is a fluent representing the missing activity. Let us leave aside how this simplified semantic representation of the VP is combined with the agent and the following NP. What is important here is that f' has to be unified with a fluent provided by context or retrieved from memory. In this case, f' can be unified with two (and possibly more) activity fluents, namely *reading* or *writing*. The amount of competition between these two binding options may affect the time it takes to establish a unification link and, if our hypothesis is correct, may modulate the amplitude of the N400 elicited by (37b) compared to (37a), in which the activity *reading* is made explicit by the VP itself. Preliminary results indicate that the word 'finished' in (37b) elicits a larger N400 compared to 'read' in (37a).[30]

4.2.2. *Sense and reference as algorithm and value.* We now turn to the second ingredient of our account of semantic unification. Semantics in the tradition of Frege makes a distinction between the meaning of an expression, the structures in which it can be interpreted, and the ideas or representations associated with it. According to Frege (34), the different ideas that a painter, a horseman and a zoologist connect with 'Bucephalus' are something quite different from (and do not contribute to determining) the sense of the word (the thought it expresses) or its reference (Alexander the Great's horse). The sense and reference of an expression are objective entities, which belong to the domain of thought and the real world respectively, while mental representations are subjective and refractory to description and comparison. In modern logical semantics, expressions are interpreted directly into the real world, which is usually represented by set-theoretic structures. The situation is different in cognitive semantics, where the relation between an expression and its domain of reference

[30] We refer the reader to the work of Choma (14) for details on the formal analysis and for EEG/ERP results.

is broken down into two: a properly semantic relation between an expression and a mental representation, and a causal relation between the cognitive model and the real world. From a different perspective, DRT views interpretation as the construction of a cognitive model, intermediate between the syntactic representation and the world.

Although realistic semantics in Frege's tradition and cognitive semantics seem radically incompatible, there is no principled opposition between the two approaches (50); indeed, it all depends on the way we choose to model Frege's notions of sense and reference. Suppose that, following Thomason (105), Muskens (77) and Moschovakis (75), we equate the sense of an expression with an algorithm that computes its denotation. We can assume that senses of expressions are stored as algorithms in semantic memory, an idea which has been applied to model the meaning of verb phrases in (114). Notice that Frege's concerns with the instability of mental representations do not apply to our proposal. First, mental imagery as evoked by linguistic expressions, if at all relevant for determining meaning, is regimented by Hintikka's method, which is aimed precisely at revealing the systematic properties of the objects of intuition; thus, against Frege, one could argue that mental representations are not completely and not always idiosyncratic. Second, algorithms are not accessible through introspection, but need to be posited by the theorist and described at an independent level of analysis. Algorithms can be viewed as hypotheses or theoretical constructs, and as such are rather different from the kind of entities to which Frege's qualms would apply.

We now turn to meaning at the discourse level. Discourse representations have to have particular properties in order to fit with a computationally viable notion of interpretation (50; 114; 101). To illustrate this point, consider the following sentences:

(39) a. John was crossing the street when he was hit by a truck.
b. John was crossing the street when he was stung by a mosquito.

In (39a) John did not reach the other side of the street because the truck prevented him from doing so. In (39b) however we have no positive information on whether the mosquito sting had a terminating effect on the crossing activity. It seems unlikely that statements specifying all the possible effects and non-effects of a given event are stored in semantic memory, for performing database searches and model checking on such a large set of statements (think of all the possible non-effects of a mosquito sting) would lead to a combinatorial explosion.

What seems more plausible is that the brain treats information which cannot be derived from the discourse or world knowledge as false, until evidence for the contrary is provided. Closed world reasoning thus leads to models which are in a sense *minimal*: only information provided by discourse and world knowledge, and its logical consequences, is represented in the model. This notion has been applied to the analysis of several linguistic phenomena (50; 114) and to a revisitation of well-known observations in the psychology of reasoning (101). Minimal models are appealing because (a) the theory implies the existence a unique minimal model interpreting the discourse (114), (b) the algorithm for obtaining minimal models from the sense of the expressions involved is fully computable, (c) minimal models can be viewed as stable states of associated neural networks (101; 102).

4.2.3. *Non-monotonic inference.* Let us turn to the third ingredient. Many psycholinguistic models (69; 55) assume that lexico-semantic integration proceeds incrementally as words are encountered. We hypothesize that the human brain integrates incrementally and very rapidly not only lexical specifications or 'core meanings' but also temporal and aspectual information carried by verbs and verb phrases, which we chose to represent in the form of integrity constraints, as well as causal knowledge stored in memory. To make the point of incrementality and immediacy clear, consider examples (39a-b). Processing the progressive clause 'John was crossing the street' leads to the introduction of an integrity constraint stating that every update of the unification space should be consistent with the information that John's crossing activity was taking place in the past (114, pp. 156-160):

(40) $?HoldsAt(cross, t), t < now$ `succeeds`

where the fluent f denotes the crossing activity. Up to this stage, there is no evidence that John encountered an obstacle while crossing the street. The system will now exploit the principle of inertia which, as we have already seen, states that if an event e has initiated an activity f at time t and nothing happens to stop the activity between t and t', then f is still going on at t'. For VPs which describe a goal-directed or telic activity, the consequence is that, if no such obstacle occurs, the activity will lead to the attainment of the goal state. The initial minimal model, computed while processing the progressive clause, represents the crossing activity as successful, that is John reached the other side of the street (see (114, pp. 159-161) for a rigorous statement and a computational proof). This minimal model is an *estimate* of the meaning of the progressive clause, and as such can be rejected on the basis of subsequent discourse. The minimal model is simply extended when the subordinate clause in (39b) is processed, since no positive evidence for the occurrence of obstacles is provided. The inference that John eventually reached the other side of the street can still be drawn. Notice that, without closed world reasoning, that is, in a model built upon classical logic, semantic binding operations would soon reach a dead end: if the system could not assume that the mosquito sting had no effect on the crossing activity, then simply nothing would follow about John having reached the other side of the street, which is not what is implied by these sentences. So immediacy tells us that the inference that John reached the other side of street is drawn as soon as the progressive clause is processed, and that this holds for (39a) as well as for (39b) (2). An important consequence of this is that minimal models may have to be recomputed on-line when (previously missing) positive information is provided by discourse, as in (39a). 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, the information that John was hit by a truck will force the system to withdraw the former inference. Inference on minimal models is thus non-monotonic, that is, it can be canceled whenever evidence for the contrary is provided.

5. Conclusion

Throughout the chapter we have been claiming that the interactions between linguistics, psycholinguistics and the cognitive neuroscience of language have often been occasional and methodologically unconstrained. Applying Marr's scheme to language, we have seen what a more articulated interplay of these disciplines could look like. The reader might however feel that we have misrepresented the status of the relations between linguistic theory and cognitive science. In fact, one might argue, there have been various attempts to determine, for instance, whether grammars can be used to predict and explain complexity effects in sentence processing experiments. One of the most important breakthroughs of formal grammars into psycholinguistics was achieved in the late 1960s with the approach called Derivational Theory of Complexity (DTC) (32), which was supposed to establish the dependence of the complexity of sentence processing, as revealed for instance by reading times, on the complexity of the derivation of the syntactic structure of the sentence within a formal grammar. Most commentators agree that the research programme failed and was abandoned after the controversial results obtained by Slobin (97) and Forster & Olbrei (33) among others: while the levels of representation of the grammar (deep structure, surface structure, logical form, etc.) might be real, transformations are not (107). Nevertheless, the practice of invoking operations postulated by linguistic theory to explain empirical data is still in vogue today (see for example (40)). Since there have been and there are interactions between linguistics and cognition, as exemplified by DTC and related approaches, why should we recommend something different?

One way of broaching the issue is to ask what lesson can be learned from attempts of using a theory of competence to account for language processing data. The failure of DTC, in particular, can be interpreted in three ways. The first is to say that the theory has been falsified by the observations: since the grammar predicted an increased processing load for sentences requiring more transformations, and such processing effects were not observed, the grammar is incorrect. The second (85) is that there is only a weak connection between the grammar and processing mechanisms: because a correlation of transformational and processing complexity was expected on the basis of the assumption that the operations posited by the grammar approximate the operations carried out by the physical system, the failure to observe a correlation implies that the assumption is incorrect. The third is that competence theories cannot be directly tested on processing data: a cognitive model is able to account for the behavior of an information processing system to the extent that it correctly characterizes the computational theory *and* the relevant algorithms and memory mechanisms; therefore, a failure to take into account Marr's intermediate level of analysis may lead to discrepancies between theory and observation.

The claim that transformational grammar had been falsified by experimental results was never held into serious consideration. The second conclusion was probably the most widely accepted. As Jackendoff has observed (56), the failure of DTC contributed to hardening the divide between competence and performance: following a line of reasoning which we have encountered in section 3.1, the controversial results were eventually taken as irrelevant for transformational grammar, and vice versa. One can only agree, if this implies that the data may become relevant for linguistics when the grammar is subsumed under an appropriate methodological framework such as Marr's. But the conclusion amplified by Jackendoff is only a starting point toward identifying and understanding the causes of the failure of DTC. In this respect, we find the third reconstruction more convincing: the research programme failed because the algorithms and memory mechanisms supporting the grammar had not been taken into account. The picture of relations between levels of analysis sketched here may sound rather unsettling, for it makes the task of providing a comprehensive theory of language processing extremely demanding, if not practically impossible. However, even if that was true, the enterprise would be still worth pursuing, as it may eventually help us to refine our criteria of what counts as a truly comprehensive view of the language faculty.

Acknowledgments. We wish to thank Oliver Bott, Travis Choma, Bart Geurts, Fritz Hamm, Karl Magnus Petersson, Martin Stokhof, Julia Uddén, Theo Vosse and Roel Willems for their useful comments on the problems discussed here. We are grateful to NWO (the Netherlands Organization for Scientific Research) for support under grant NWOCOG/04-19.

References

[1] L. Åqvist, F. Guenthner, and C. Rohrer. Definability in ITL of some subordinate temporal conjunctions in english. In F. Guenthner and C. Rohrer, editors, *Studies in Formal Semantics*, pages 201–221. North-Holland, 1978.

[2] G. Baggio and M. van Lambalgen. The processing consequences of the imperfective paradox. *Journal of Semantics*, 24:307–330, 2007.

[3] R.A. Berman and D.I. Slobin (eds.). *Relating events in narrative*. Lawrence Erlbaum Associates, 1994.

[4] D. Bolinger. Judgments of grammaticality. *Lingua*, 21:34–40, 1968.

[5] G. Boole. *An Investigation of the Laws of Thought*. Dover, 1958.

[6] J.D. Bransford, J.R. Barclay, and J.J. Franks. Sentence memory: A constructive versus an interpretive approach. *Cognitive Psychology*, 3:193–209, 1972.

[7] C.M. Brown and R. Hagoort (eds.). *The Neurocognition of Language*. Oxford University Press, New York, 1999.

[8] M. Bunge. Philosophical problems in linguistics. *Erkenntnis*, 21:107–173, 1984.

[9] T. Burge. *Truth, Thought, Reason. Essays on Frege*. Clarendon Press, 2005.

[10] M.N. Carminati, L. Frazier, and K. Rayner. Bound variables and c-command. *Journal of Semantics*, 19:1–34, 2000.

[11] P. Carruthers. *Language, Thought and Consciousness*. Cambridge University Press, 1996.

[12] P. Carruthers. The cognitive functions of language. *Behavioral and Brain Sciences*, 25:657–725, 2002.

[13] L.L. Cavalli-Sforza and M.W. Feldman. *Cultural transmission and evolution: a quantitative approach*. Princeton University Press, 1981.

[14] T. Choma. Coercion and compositionality: An EEG study. Master's thesis, University of Amsterdam, 2006.

[15] N. Chomsky. *The Logical Structure of Linguistic Theory*. MIT Press, 1955.

[16] N. Chomsky. *Syntactic Structures*. De Gruyter-Mouton, 1957.

[17] N. Chomsky. *Aspects of the Theory of Syntax*. MIT Press, 1965.

[18] N. Chomsky. *Lectures on Government and Binding*. Foris, 1981.

[19] N. Chomsky. *The Minimalist Program*. MIT Press, 1995.

[20] N. Chomsky, A. Belletti, and L. Rizzi. *On Nature and Language*. Cambridge University Press, 2002.

[21] B. Comrie. *Aspect*. Cambridge University Press, Cambridge, 1976.

[22] B. Comrie. *Tense*. Cambridge University Press, Cambridge, 1985.

[23] M. Cresswell. Interval semantics and logical words. In C. Rohrer, editor, *On the Logical Aspects of Tense and Aspect*, pages 7–29. Narr, 1977.

[24] W. Croft and D.A. Cruse. *Cognitive Linguistics*. Cambridge University Press, 2004.

[25] M. Devitt. Intuitions in linguistics. *British Journal of Philosophy of Science*, 57:481–513, 2006.

[26] 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.

[27] T.A. Erickson and M.E. Matteson. From words to meanings: A semantic illusion. *Journal of Verbal Learning and Verbal Behaviour*, 20:540–552, 1981.

[28] 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.

[29] J.A. Feldman and D.H. Ballard. Connectionist models and their properties. *Cognitive Science*, pages 205–254, 1982.

[30] W.T. Fitch and M.D. Hauser. Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656):377–380, 2004.

[31] W.T. Fitch, M.D. Hauser, and N. Chomsky. The evolution of the language faculty: clarifications and implications. *Cognition*, 97(2):179–210; discussion 211–225, 2005.

[32] T.G Fodor, J.A. Bever and M.F. Garrett. *The Psychology of Language*. McGraw Hill, 1974.

[33] K. Forster and I. Olbrei. Semantic heuristics and syntactic analysis. *Cognition*, 2(3):319–347, 1973.

[34] G. Frege. On sense and reference. In P. Geach and M. Black, editors, *Translation from the Philosophical Writings of Gottlob Frege*, pages 56–78. Blackwell, 1980.

[35] T.Q. Gentner, K.M. Fenn, Daniel M., and H.C. Nusbaum. Recursive syntactic pattern learning by songbirds. *Nature*, 440(7088):1204–7, 2006.

[36] B. Geurts and F. van der Slik. Monotonicity and processing load. *Journal of Semantics*, 22:97–117, 2005.

[37] H. Geurts. *Executive functioning profiles in ADHD and HFA*. PhD thesis, Amsterdam Free University, 2003.

[38] E.M. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[39] 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.

[40] Y. Grodzinsky. The neurology of syntax: Language use without broca's area. *Behavioral and Brain Sciences*, 23(1):1–21; discussion 21–71, 2000.

[41] 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, New York, 1998.

[42] P. Hagoort. How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage*, 20:S18–S29, 2003.

[43] P. Hagoort. On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9:416–423, 2005.

[44] 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.

[45] P. Hagoort. The Memory, Unification, and Control (MUC) model of language. In A.S. Meyer, L.R. Wheeldon, and A. Krott, editors, *Automaticity and Control in Language Processing*, pages 243–270. Psychology Press, Philadelphia, 2006.

[46] 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, Hillsdale, NJ, 1994.

[47] 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.

[48] 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.

[49] P. Hagoort, L. Hald, M. Bastiaansen, and K.M. Petersson. Integration of word meaning and world knowledge in language comprehension. *Science*, 304:438–441, 2004.

[50] F. Hamm, H. Kamp, and M. van Lambalgen. There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 22:1–40, 2006.

[51] M.D. Hauser, N. Chomsky, and W.T. Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579, 2002.

[52] J. Hintikka. The emperor's new intuitions. *Journal of Philosophy*, 96:127–147, 1999.

[53] J. R. Hurford. The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, 26:261–316, 2003.

[54] R. Jackendoff. On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26:89–114, 1987.

[55] R. Jackendoff. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford, 2002.

[56] R. Jackendoff. Precis of Foundations of Language: Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*, 26:651–65; discussion 666–707, 2003.

[57] P.N. Johnson-Laird. Mental models in cognitive science. *Cognitive Science*, 4:71–115, 1980.

[58] W. Kohler. *The mentality of apes*. Harcourt Brace and World, New York, 1925.

[59] M. Kutas and S.A. Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207:203–205, 1980.

[60] M. Kutas and S.A. Hillyard. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163, 1984.

[61] W. Labov. When intuitions fail. *Papers from the Parasession on Theory and Data in Linguistics. Chicago Linguistic Society*, 32:77–106, 1996.

[62] W. Levelt. Some psychological aspects of linguistic data. *Linguistische Berichte*, 17:18–30, 1972.

[63] D. Lewis. General semantics. *Synthese*, 22:18–67, 1970.

[64] S. Luck. *An Introduction to the Event-Related Potential Technique*. MIT Press, Cambridge, MA, 2005.

[65] A.R. Luria. *Cognitive Development: Its Social and Cultural Foundations*. Harvard University Press, Cambridge, MA, 1976.

[66] A. Marantz. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22:429–445, 2005.

[67] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, 1982.

[68] D. Marr and T. Poggio. From understanding computation to understanding neural circuitry. *MIT AI Memos*, 357:1–22, 1976.

[69] W. Marslen-Wilson. Access and integration: Projecting sound onto meaning. In W. Marslen-Wilson, editor, *Lexical representation and process*, pages 3–24. MIT Press, Cambridge, MA, 1989.

[70] 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 Massachussets, Amherst, 1994.

[71] 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.

[72] R. McKinnon and L. Osterhout. Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes*, 11:495–523, 1996.

[73] C.T. McMillan, R. Clark, P. Moore, C. Devita, and M. Grossman. Neural basis of generalized quantifier comprehension. *Neuropsychologia*, 43:1729–1737, 2005.

[74] M. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall, 1967.

[75] Y. Moschovakis. Sense and denotation as algorithm and value. In J. Oikkonen and J. Vaananen, editors, *Lecture Notes in Logic, 2*. Springer, 1994.

[76] T.F. Münte, K. Schiltz, and M. Kutas. When temporal terms belie conceptual order. *Nature*, 395:71–3, 1998.

[77] R. Muskens. Sense and the computation of reference. *Linguistics and Philosophy*, 28(4):473–504, 2005.

[78] 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.

[79] M.A. Nowak and N.L. Komarova. Towards an evolutionary theory of language. *Trends in Cognitive Sciences*, 5(7):288–295, 2001.

[80] L. Osterhout and P. Holcomb. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31:785–806, 1992.

[81] A. Özyürek, R.M. Willems, S. Kita, and P. Hagoort. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 19:605–616, 2007.

[82] P. Perruchet and A. Rey. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin and Review*, 12(2):307–313, 2005.

[83] K.M. Petersson. *Learning and Memory in the Human Brain*. PhD thesis, Department of Clinical Neuroscience, Karolinska Institutet Stockholm, 2005.

[84] K.M. Petersson. On the relevance of the neurobiological analogue of the finite-state architecture. *Neurocomputing*, 65-66:825–832, 2005.

[85] C. Phillips. *Order and Structure*. PhD thesis, Massachusetts Institute of Technology, 1996.

[86] S. Pinker and R. Jackendoff. The faculty of language: what's special about it? *Cognition*, 95(2):201–236, 2005.

[87] F. Pulvermuller. Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6:576–582, 2005.

[88] 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.

[89] 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.

[90] 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.

[91] W. V. Quine. Methodological reflections on current linguistic theory. *Synthese*, 21:386–398, 1970.

[92] M. Riesenhuber and T. Poggio. Are cortical models really bound by the 'binding problem'? *Neuron*, 24:87–93, 1999.

[93] M. D. Root. Speaker intuitions. *Philosophical Studies*, 29:221–234, 1976.

[94] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, New York, 1962.

[95] S. Scribner. *Mind and social practice*. Cambridge University Press, Cambridge, 1997.

[96] M. Singer. Discourse inference processes. In M.A. Gernsbacher, editor, *Handbook of Psycholinguistics*. Academic Press, 1994.

[97] D. Slobin. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behaviour*, 5:219–227, 1966.

[98] M.J. Spivey and Gonzalez-Marquez M. Rescuing generative linguistics: Too little, too late? *Behavioral and Brain Sciences*, 26:690–691, 2003.

[99] M. Steedman. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 25:723–753, 2002.

[100] K. Stenning. How did we get here? A question about human cognitive evolution. Frijda Lecture, University of Amsterdam, 2003.

[101] K. Stenning and M. van Lambalgen. Semantic interpretation as computation in non-monotonic logic: The real meaning of the suppression task. *Cognitive Science*, 29:919–960, 2005.

[102] K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science*. MIT Press, Cambridge, MA, 2007.

[103] P. Suppes. Axiomatic methods in science. In M.E. Carvallo, editor, *Nature, Cognition and System II*, pages 205–232. Kluwer Academic Publishers, 1992.

[104] P. Suppes. *Representation and Invariance of Scientific Structures*. Lecture Notes. Center for the Study of Language and Information, 2001.

[105] R. Thomason. A model theory for propositional attitudes. *Linguistics and Philosophy*, 4:47–70, 1980.

[106] M. Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, 1999.

[107] D.J. Townsend and T.G. Bever. *Sentence Comprehension. The Iintegration of Habits and Rules*. MIT Press, 2001.

[108] 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.

[109] 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.

[110] J.J. van Berkum, C.M. Brown, and P. Hagoort. Early referential contex effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41:147–182, 1999.

[111] 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.

[112] J.J. 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.

[113] F. van der Velde. Is the brain an effective turing machine or a finite-state machine? *Psychological Research*, 55:71–79, 1993.

[114] M. van Lambalgen and F. Hamm. *The Proper Treatment of Events*. Blackwell, Oxford, 2004.

[115] C. von der Malsburg. The correlation theory of brain function. Internal Report 81-2, Dept. of Neurobiology, Max- Planck-Institute for Biophysical Chemistry, Goettingen, Germany, 1981. Reprinted in: E. Domany, J.L. van Hemmen and K. Schulten (eds.) *Models of neural networks II*, Springer Verlag, 1994.

[116] C. von der Malsburg. The what and why of binding: The modeler's perspective. *Neuron*, 24:95–104, 1999.

[117] 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.

[118] P. C. Wason. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20:273–281, 1968.

[119] V.H. Yngve. A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104:444–466, 1960.