

THE PROCESSING CONSEQUENCES OF COMPOSITIONALITY

GIOSUÈ BAGGIO, MICHEL VAN LAMBALGEN, AND PETER HAGOORT

Author affiliations:

GIOSUÈ BAGGIO
SISSA International School for Advanced Studies, Trieste

MICHEL VAN LAMBALGEN
Institute for Logic, Language and Computation
University of Amsterdam

PETER HAGOORT
Donders Centre for Cognitive Neuroimaging
Radboud University Nijmegen *and*
Max Planck Institute for Psycholinguistics, Nijmegen

Address correspondence to:

GIOSUÈ BAGGIO
SISSA International School for Advanced Studies
Cognitive Neuroscience Sector
via Bonomea 265
34136 Trieste – Italy
E-mail: gbaggio@sissa.it

1. INTRODUCTION

It is often argued that the principle of compositionality is formally vacuous: any grammar can be given a compositional semantics (Janssen 1986; Zadrozny 1994),¹ which implies the principle is also empirically vacuous: if a compositional analysis of any linguistic structure
 5 can be given, compositionality is always upheld by the data. To be sure, the meaning of any complex expression can be viewed as a function of the meanings of its constituents and the syntactic mode of combination, *provided enough complexity is built into the structures involved*, that is the lexicon and the syntax. These are not motivated on independent grounds, as their characterization serves the sole purpose of yielding a compositional theory (Groenendijk &
 10 Stokhof 1991).

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

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

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

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

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

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

That this is not necessarily so can be illustrated using an example from Keenan (1979). In an adjective–noun construction, the noun is the argument fed into the adjective, which is viewed as a function. Keenan observes that the interpretation of the function word seems to be determined by its argument: compare for instance the different meanings of the adjective 'flat' in 'flat tyre', 'flat beer', 'flat note' etc. It is of course technically possible, as Keenan notes, to replace the single function 'flat' by a disjunctively defined function, where each of the disjuncts corresponds to a separate meaning for 'flat', with suitable selection restrictions on the argument. However, this technical solution is surely paradoxical: compositionality was invoked to account for productivity, which seemed hard to explain in terms of storage only; but, in this case, compositionality can apparently be salvaged only by *increasing* the demand on storage! From a processing perspective, it would be much better if there were a single computational mechanism generating the meaning of a *flat* + *noun* construction, starting from a single basic meaning of 'flat'. These considerations show that the principle of compositionality is affected by its ambiguous status: as a formal desideratum on the one hand, and a processing hypothesis on the other.

1.2. The systematicity argument. A second argument in favor of compositionality is based on the observation that languages are systematic, that is, the ability to understand certain utterances is connected to the ability to understand certain others. For instance, any native speaker of English that understands 'John loves Mary' also understands 'Mary loves John'.

SYSTEMATICITY *Anyone who understands the complex expressions e and e' built up through the syntactic operation F from the constituents e_1, \dots, e_n and e'_1, \dots, e'_n respectively, can thereby also understand any other expression e'' built up through*
 70 *F from expressions among $e_1, \dots, e_n, e'_1, \dots, e'_n$.*

Systematicity seems to entail compositionality, but the issue here is whether languages are systematic in the sense above. For instance, anyone who understands 'The dog is asleep' and 'The cat is awake' can certainly understand 'The dog is awake' and 'The cat is asleep'. However, would everyone who understands 'within an hour' and 'without a watch' also
 75 understand 'within a watch' and 'without an hour'? The definition presupposes that e'' is a meaningful expression and that $e_1, \dots, e_n, e'_1, \dots, e'_n$ can be freely combined and substituted while keeping F constant. But the fact that we can hardly make sense of 'within a watch' and 'without an hour' suggests that this is not the case, thus languages are not systematic in the sense of the definition above. Now, let us suppose for a moment that this difficulty
 80 can be overcome. Would then systematicity force compositionality upon us? It seems not, for systematicity says that, given the sentences 'The dog is asleep' and 'The cat is awake', from the meanings of 'the dog', 'the cat', 'is asleep' and 'is awake' plus the syntax, one is able to understand the meaning of 'The dog is awake'. Compositionality makes however a stronger claim, namely that the meanings of 'the dog' and 'is awake' plus the syntax are
 85 sufficient to do that. So, even if systematicity held, it would not buy compositionality (see Johnson (2004), Szabó (2007), Pullum & Scholz (2007) for a discussion and further issues).

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

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

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

95 The trouble with the methodological argument is that compositionality is highly theory dependent (Partee et al. 1990). Ideally, when looking for an explanation of a given linguistic phenomenon, one takes the syntax and the semantics to be fully specified formal systems. It is then a definite question whether that phenomenon allows for a compositional treatment. If it does not, one may take this as a cue for changing the semantics. In practice, however,

100 the explanation of a new phenomenon of interest often leads to changes in *both* syntax and semantics. Compositionality then becomes a soft constraint indeed.

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

110 **1.4. The modularity argument.** A fourth argument leading to compositionality is suggested by a view of the language faculty as a ‘cognitive module’. Fodor (1983) lists nine properties characterizing modular systems: domain specificity, fast and mandatory operation, limited central access to modular representations, informational encapsulation, shallow outputs, fixed neural architecture, specific breakdown patterns, and characteristic ontogenetic pace and sequencing. Of these, the most relevant for our purposes is *informational encapsulation*. This is the idea that perceptual systems – language included – are relatively impenetrable to the bulk of the knowledge internally represented by the organism. Informational encapsulation says that there are tight constraints on the flow and handling of extra-modular information *within* the module *prior to* the production of an output.

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

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

representation of the clause.³ Compositionality would then constrain the kind of traffic that can occur between the two modules.

135 So regardless of one's choice of modular architecture, compositionality remains relevant insofar as it acts as a counterpart of information encapsulation at the level of the description of linguistic structure. The link between compositionality and informational encapsulation can be made more explicit: if the composition of meanings is not affected by extra-modular knowledge, then one can characterize the meaning of any complex expression as a function
140 of the meanings of its constituents and syntactic rules *only*, all of which are readily available within the module.⁴ Clearly, this hinges very much on what one assumes is contained in the module(s) – and this will be a recurrent theme in this chapter. What bears some emphasis here is that the degree to which a system is informationally encapsulated can be determined only based on empirical data. Hence, with this 'argument from modularity' in place, it also
145 becomes possible to treat compositionality as a processing principle, that is, as a constraint on the range of data structures involved in language processing.

2. COMPOSITIONALITY AS A PROCESSING PRINCIPLE

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

INCREMENTAL COMPOSITION *The meaning of a complex expression at some processing stage σ is computed based on the constituent expressions processed at
155 σ and of the syntactic structure built up at σ .*⁵

This definition is silent as to whether meaning assembly involves the lexicon and the syntax *only*, or whether other sources of information could enter the composition process. For this we need another definition, which can be combined with the former:

SIMPLE COMPOSITION (i) *The meanings of elementary expressions are the only
160 constraints on content in the computation of the meaning of a complex expression.*

³For a modular view of semantics, see among others Borg (2004) and Robbins (2007) for a critical discussion.

⁴Jackendoff (1997) makes the same point: "The hypothesis of syntactically transparent semantic composition has the virtue of theoretical elegance and constraint. Its effect is to enable researchers to isolate the language capacity – including its contribution to semantics – from the rest of the mind, as befits a modular conception."

⁵For simplicity, we may assume that each word corresponds to a processing stage, and vice versa. An important theoretical question is whether assuming finer-grained processing steps would lead to local inconsistencies between incrementality and compositionality.

(ii) *The syntax of elementary expressions is the only constraint on structure in the computation of the meaning of a complex expression.*

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

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

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

INSIDE-OUT PRINCIPLE *The proper direction of a semantic analysis is from the inside out – that is, from the bottom up, or from elementary to complex meanings.*

Together, these consequences suggest that we ought to take the notion of a ‘function’ in the formulation of the principle very seriously. Semantic computation of complex expressions is function application, so *the meanings of simple expressions are not changed by virtue of the fact that they occur as arguments of the function.*⁶ This shows that the principle of compositionality is tied to one very particular form of linguistic productivity, exemplified by the usual rules: if S is a sentence, one can form a new sentence ‘I think that S ’; if S_1, S_2 are sentences, one can form ‘ S_1 and S_2 ’; etc. But there exist other forms of productivity in natural languages, which do not have this function-like character. An example is the progressive construction in English when applied to stative verbs (Croft 1998):

- (2) a. She resembles her mother.
b. *She is resembling her mother.

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

- c. She is resembling her mother more and more every day.

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

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

If recomputation were to be supported, where would that leave compositionality and the work it is supposed to do? The strict reading as embodied in Hintikka’s principles presents semantic composition as entirely inside-out/bottom-up. The recomputation account is also partially outside-in/top-down. In theory, this has the consequence that the meaning which is assigned to an expression is always provisional, a situation that cannot occur on a literal reading of compositionality (but see 3.1). However, there is room for both accounts, because the productivity of language is a two-dimensional phenomenon. On the vertical dimension, there is productivity due to increased syntactic complexity; here simple composition has an important role to play. There is however also a horizontal dimension to productivity – here it is not so much the syntactic complexity that increases, but new meanings of a given clause are produced by varying syntactic or semantic aspects of other clauses. Thus, if we replace the adverbial ‘more and more every day’ in (2-c) with ‘in broad daylight’, the progressive is no longer felicitous, because the verb ‘resemble’ then has its default stative meaning. The

horizontal dimension of productivity seems to call for some form of enriched composition, that is, ways of combining meaning that allow top-down influences and recomputation.

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

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

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

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

The *Fregean value* of e is the equivalence class of e modulo \sim . Hodges shows that taking the Fregean value of e as its meaning yields a compositional semantics. Therefore, if we assume that the module contains all Fregean values, modularity is restored. Moreover, this notion of meaning is pleasingly context-sensitive. Hodges gives the example of sentences which have different social connotations, for instance ‘he is intoxicated’ vs. ‘he is pissed’ (Hodges 2006). The contexts in which these sentences are acceptable are very different. In this sense, these two sentences have different meanings in Hodges’ semantics, whereas they would be treated as synonymous with ‘he is drunk’ in a more standard framework.

Although Hodges’ proposal formally restores compositionality, it does so in a way that upsets the balance between storage and computation, and renders it unclear how meanings can be acquired. The Fregean value of e is defined by means of all possible uses of e , and it is doubtful that these are available to a young language learner. In other words, learning meanings is a gradual process, so Fregean values should be partial objects, in the sense of being defined by means of a subset of all possible uses of e , and being subject to update and revision as learning proceeds, which is not what Hodges suggests. The point can be amplified by considering what it means to *know* the Fregean value of e . There seem to be two components to this:

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

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

3. EXPERIMENTAL DATA BEARING ON COMPOSITIONALITY

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

Relevant to this issue is the well-known ‘Moses illusion’ (Erickson & Matteson 1981). When asked ‘How many animals of each sort did Moses put on the ark?’, subjects tend to respond ‘two’ without questioning the (false) presupposition that Moses was the biblical character who did that. Similar results have been obtained with questions such as ‘After an air-crash, where should the survivors be buried?’ (Barton & Sanford 1993) or ‘Can a man

marry his widow's sister?' (Sanford 2002). Hearers seem to be processing these sentences superficially enough to miss the false presuppositions. Ferreira et al. (2002) and Ferreira & Patson (2007) suggest that these data challenge compositionality. Consider the Moses question. It would seem that the meaning computed by hearers and the meaning derived compositionally are in some important respect different. If the former were a function of the meanings of the constituents and the syntax, 'Moses' would mean Moses and hearers would notice the false presupposition. This seems an instance of a non-compositional process.

One might argue against this conclusion by emphasizing that these data just show that each word in a sentence does not contribute its full meaning (Sanford & Sturt 2002). The 'full meaning' of 'Moses' need not be retrieved, only some feature made salient by the context (Sanford & Sturt 2002; Sanford 2002). This might be 'biblical character', 'patriarch' or some other feature responsible for the semantic proximity of 'Noah' and 'Moses' (van Oostendorp & de Mul 1990). Feature sharing is thus what gives rise to the illusion. Nonetheless, the fact that the lexicon may be a locus of shallow processing (or retrieval, as the case may be) does not speak against compositionality. Simple composition entails that the lexicon is the only provider of content for complex meanings, though not that *full* lexical representations must be used. The latter seems too strong a requirement to press upon either simple or enriched composition. For if there are such entities as 'full lexical meanings' – and there are reasons to be skeptical about that – they can hardly be used on most processing tasks, because of the massive amount of information, presumably continuous with non-lexical knowledge, that would be fed into the composition process. Incremental, context-sensitive feature selection during the retrieval or word meaning would thus be the default. Semantic illusions would become a special case, in which some critical semantic feature is shared between the word triggering a true presupposition ('Noah') and the word triggering a false presupposition ('Moses'). In summary, compositionality can be refined to accommodate semantic illusions, by allowing composition to make use of partial lexico-semantic representations.

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

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

The 'garden-path model' (Frazier 1987) hypothesizes that 'the baby' is initially parsed and interpreted as the direct object of 'dressed'. Only when the verb 'played' is encountered the syntactic and semantic representations are revised to the effect that 'the baby' is the subject of 'played'. One question here is whether the initial representation, featuring 'the baby' as a direct object, is at all maintained in memory. Christianson et al. (2001) show that, while

330 readers correctly respond for the affirmative to 'Did the baby play in the crib?', they also
give a positive answer to 'Did Anna dress the baby?'. No grammar or parser on the market
would allow for the same token NP to play two functional roles, subject and direct object.
And yet, this appears to be precisely the interpretation computed by readers. Ferreira et al.
(2002) take this as "clear evidence that the meaning people obtain for a sentence is often not
335 a reflection of its true content" – that is, it is not built up compositionally.

Does the existence of 'lingering misinterpretations' demonstrate that the processing of
garden-path sentences is non-compositional, as suggested by Ferreira and colleagues? There
are at least two ways of accounting for the data, from which different answers ensue. On
one account, the last interpretation subjects come up with is that while Anna dressed, the
340 baby played in the crib, which corresponds to the revised parse whereby 'the baby' is the
subject of 'played', and is no longer the direct object of 'dressed'. Interpretation is therefore
non-monotonic, that is, allowing for revisions on earlier structures. This however requires
a refinement of compositionality that was introduced earlier on – incremental composition.
On this view, both the initial and the revised interpretations can be derived compositionally,
345 and simple composition seems enough in this case. The persisting misinterpretation would
be rather an effect of memory architecture or neural implementation. One aspect of the data
of Christianson et al. (2001) which supports this story is that misinterpretations are more
frequent when the head of the misanalyzed phrase occurs early. That is, misinterpretations
are more likely to persist the longer they have been part of the initial discourse model.

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

3.3. Interlude: event-related brain potentials (ERPs). One aspect of enriched composition
is that the stored meanings of elementary expressions are not the only source of content for

365 complex meanings. It can be suggested that this challenges informational encapsulation, if
one can demonstrate that such additional semantic information is handled by the module
(or component) before an output is produced (Fodor 1983). Some experimental techniques
in cognitive neuroscience allow one to make timing inferences, the most sensitive and direct
of which is based on event-related brain potentials (ERPs) (Van Berkum 2004).

370 The post-synaptic currents generated by the neocortex (electroencephalography, or EEG)
can be recorded by placing a number of electrodes on the scalp, amplifying the signal and
plotting the observed voltage changes as a function of time (Luck 2005). ERPs are defined
as electrical potentials time-locked to some event of interest, such as the onset of a stimulus.
Averaging over a relatively large number of trials of the same condition is the most widely
375 used approach to obtaining ERPs.

An ERP component that is particularly relevant here is the N400. This is a negative shift
starting around 250 ms after word onset, peaking at 400 ms and lasting until approximately
550 ms. Every content word elicits an N400, but the amplitude of the component, relative to
a control condition, is dependent upon the degree of semantic relatedness of the given word
380 with its sentence (Kutas & Hillyard 1980; Kutas & Hillyard 1984; Hagoort & Brown 1994) or
discourse context (van Berkum et al. 1999; van Berkum et al. 2003). There is evidence that
the N400 does not just reflect lexical access, but the *integration* of a word's meaning into the
unfolding semantic representation (see Brown & Hagoort (1993) and Li et al. (2008) among
others for experimental data and Hagoort et al. (2009) for a discussion and a comparison
385 with alternative accounts of the N400). The N400 can be seen as an index of the complexity
of initial attempts to combine the meaning of the given word with the meanings of the
expressions already processed.

Another ERP effect that will be of interest here is the P600. This is a positive shift starting
around 500 ms following the onset of the word and lasting for about 500 ms (Osterhout &
390 Holcomb 1992; Hagoort et al. 1993). Larger P600 effects are elicited by sentences containing
violations of syntactic constraints (such as phrase structure, subcategorization, agreement),
temporarily syntactically ambiguous sentences, garden-path sentences, and constructions
which show high syntactic complexity (Hagoort et al. 2001; Friederici 2002; Hagoort 2003).
In relation to incremental composition, we can regard the P600 as an index of the time and
395 resources involved in attaching a given word in the syntactic representation computed thus
far (Hagoort 2003). How do N400 and P600 data bear on compositionality?

3.4. World knowledge. Relevant to this question is an ERP study by Hagoort et al. (2004),
using true (4-a), false (4-b) and semantically anomalous (4-c) sentences:⁷

⁷The stimuli were in Dutch and participants were native Dutch speakers.

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

The words ‘white’ and ‘sour’ evoked very similar N400s, in both cases larger than the N400 elicited by ‘yellow’. Integrating the meanings of ‘white’ and ‘sour’ in the ongoing semantic representation is thus relatively hard. This suggests that, upon encountering ‘Dutch trains’,
 405 features are retrieved which code for the color of Dutch trains – typically yellow-blue – and are responsible for the additional processing costs associated with ‘white’.

While it is notoriously hard to define ‘core’ semantic features, separating linguistic from world knowledge, it is nevertheless possible to identify features which are *invariant* across the individuals and communities using the relevant word. That ‘sour’ cannot be applied to
 410 (Dutch) trains seems a piece of invariant knowledge, and in that sense is a fact of linguistic predication (‘linguistic knowledge’). However, trains differ in color and other properties in space and time, hence that ‘white’ cannot be applied to Dutch trains reflects a contingent state of affairs which not all users of the expressions ‘train’ or ‘Dutch train’ may be aware of (‘world knowledge’). The N400 data of Hagoort et al. show that ‘white’ is hard to integrate
 415 in the sentence context, which in turn suggests there is something in the meaning of ‘Dutch trains’ which makes integration hard. This must be knowledge that Dutch trains are yellow and blue, and not white. It thus seems that during processing meanings are computed – for instance, of the compound ‘Dutch trains’ – encompassing invariant *and* community-specific semantic information, that is, linguistic *and* world knowledge. As for compositionality, this
 420 may mean two things, depending on one’s view of the lexicon: either the lexicon includes declarative memory in its entirety, and then simple composition seems enough to account for the similarity between the N400 effects, or the lexicon includes invariant meanings only, and then enriched composition – the thesis that the lexicon is not the only source of semantic content – is necessary to explain the observed N400s.⁸

425 **3.5. Co-speech gestures.** At least another ERP study reporting modulations of the N400 seems relevant for our discussion. Özyürek et al. (2007) show that larger N400s are elicited by co-speech gestures which do not match with the semantic content of the accompanying sentence. This demonstrates that semantic information from different modalities – speech and gesture, in this case – is integrated in the same time frame. The choice mentioned above
 430 between two views of the lexicon applies here too. If schematic representations of gestures

⁸For other results bearing on the issue of world knowledge and on-line meaning composition, see Münte et al. (1998), Ferretti et al. (2007) and Baggio et al. (2008a) among others.

are stored in declarative memory, which is assumed to be entirely contained in the lexicon, simple composition seems enough to explain the data. However, there exist experimental data showing that subjects attribute different meanings to a given iconic gesture, providing evidence for the non-conventionalized nature of gesture meanings (Chauncey et al. 2004).
 435 If indeed gesture schemes are not part of the lexicon, some form of enriched composition must occur, such that the semantics of elementary expressions is just one source of content for complex meanings. This choice between two views of the lexicon, and its consequences for the status of compositionality, shows how severe is the problem of compositionality's theory-dependence, and how pressing the need for realistic constraints on the components
 440 of the grammar.

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

(5) The peanut was salted.

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

(6) The peanut was in love.

This is taken to show that discourse can override even such deep-rooted semantic features as animacy. These findings can also be read as a challenge to Hintikka's principles, for they
 455 seem to show that meaning is context-dependent, and semantic composition can proceed from the outside-in, that is, from the discourse to lexical meaning.

Processing sentences such as (6) in a fictional context might therefore involve some form of top-down composition which, if it cannot resort to the mental lexicon as a repository of contextual values of expressions (recall Hodges' argument), then it must adopt some form
 460 of enriched composition – or give up a share of informational encapsulation, which is the same. However, there seems to be another way out for compositionality. One may ask, not what changes in the meaning of 'peanut' in a context in which it is depicted as animate, but

what is preserved of the original (invariant, as we called it) meaning of the word. Animacy aside, there is no evidence that any of the other semantic features is maintained. Therefore, the word form ‘peanut’ in the fictional context considered here may just be used as a label for an animate subject or, more precisely, a proper name with a reference but no (or perhaps very little) sense. This could easily be handled in a compositional manner. Processing the adjective ‘salted’, given the plausible combination with ‘peanut’, might recover its original sense, and this would explain the larger N400. This does not detract from the interest of the data, nor from the interpretation chosen by Nieuwland & van Berkum. It does exemplify, however, the kind of problems one encounters when trying to put compositionality to test, and in particular the exceptional elbow room compositionality leaves to its application. It is precisely this resilience which has been taken by many as empirical vacuity.

3.7. Semantic attraction. A strict reading of compositionality implies that only two sorts of constraints can interact to produce complex meanings: syntax and the lexicon. A further assumption is that the syntax is an analysis of the sentence *as is given*, for instance a formal decomposition into constituents.⁹ This is all there is to the input for semantic composition. Combining lexical meanings (representations or types, depending on the ontology) based on the syntactic structure will in turn produce an input for the interpretation. In this sense, semantics is often said to be dependent on syntax. Kim & Osterhout (2005) designed an ERP study to test the extent to which syntax is actually in control of the composition process. They presented participants with sentences such as

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

and found that ‘devouring’ in (7-a) elicited a larger P600 compared to ‘devoured’ in (7-b). If the syntax were taken as is given in the sentence – in well-formed sentences, that is – and if it were only proposing input to the semantics, an N400 to ‘devouring’ should be expected: indeed (7-a) is syntactically well-formed, whereas a semantic constraint (animacy) appears to be violated. The P600 indicates that (7-a) is perceived as a syntactic violation, originating from the impossibility at ‘devouring’ of building a passive construction. At the verb’s stem ‘devour-’, the passive is the only continuation compatible with a plausible meaning for the sentence, as is testified by (7-b). The data therefore show not only that semantic attraction to a more plausible interpretation is an important source of constraints in sentence processing – which could also be concluded if ‘devouring’ induced a larger N400 – but also that such

⁹The extent to which a syntactic analysis is allowed to deviate from the surface form of an expression is a matter of considerable debate. For a discussion, see Culicover & Jackendoff (2005) and Culicover & Jackendoff (2006).

495 constraints can override syntactic cues as these are given in the input – which is what the
 P600, as a syntax-related effect, shows. Compositionality can be salvaged only by assuming
 that semantic attractors, such as ‘kids devour hearty meals’, are configurations of the lexical
 network and not, as would seem more intuitively appealing, the result of inference. But this
 move is once again paradoxical: compositionality was introduced to explain productivity,
 500 and therefore to *ease* the burden on storage; now it seems we need a growing inventory of
 stored semantic facts to maintain compositionality.

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

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

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

An interpretation of (8-a) in which the activity remains unspecified is conceivable, and
 525 it therefore falls to experimental research to provide evidence for or against the existence of
 complement coercion. A series of studies have shown that coercing sentences such as (8-a)
 result in increased processing costs compared to controls such as (8-b) (McElree et al. 2001;
 Traxler et al. 2002; Traxler et al. 2005; McElree et al. 2006; Pyllkkänen et al. 2007). This can

be taken as evidence against simple composition.¹⁰ However, a compositional analysis may
530 still be possible if the operation which is responsible for generating the enriched meaning
(say, type-shifting) is incorporated in the syntax. This choice can be criticized on different
grounds. On empirical grounds, it predicts that (8-a) would elicit a P600 (which, as we have
seen, correlates with syntactic complexity), while the available neural data reveal a different
effect than the P600 (Pykkänen et al. 2007). On theoretical grounds, a syntactic reduction
535 of complement coercion requires syntactic representations which resemble less and less a
formal decomposition into constituents (Culicover & Jackendoff 2005): the simplicity and
theoretical elegance which are gained by reintroducing compositionality are lost at the level
of syntactic structure.

4. CONCLUSIONS

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

545 Compositionality (simple composition) remains effective as an explanation of cases in
which processing complexity increases due to syntactic factors only. However, it falls short
of accounting for situations in which complexity arises from interactions with the sentence
or discourse context, perceptual cues and stored knowledge. The idea of compositionality
as a methodological principle is appealing, but imputing the complexity to one component
550 of the grammar or other, instead of enriching the notion of composition, is not always an
innocuous move, leading to fully equivalent theories. One may be tempted to believe that
equivalent theories in this sense are *in principle* indistinguishable in the face of empirical
data. However, neuroscience grants us (restricted) selective access to linguistic processes
and representations in the brain, as exemplified by the difference between N400 and P600.
555 Therefore, there is at least a chance that what appear to be neutral methodological choices
are in fact controvertible given the data. Compositionality sets also an upper bound on the
degree of informational encapsulation that can be posited by modular or component-based
theories of language: simple composition ties in with a strongly modular take on meaning
assembly, which is seen as sealed off from information streams other than the lexicon and
560 the syntax. Empirical data seem to suggest that the upper bound is not always attainable.
This implies a weakening of one's notion of compositionality, but also more complex traffic

¹⁰Evidence for enriched composition in cases of aspectual coercion has also been found. See Piñango et al. (1999)
and Piñango et al. (2006).

either within a module or between modules. So compositionality is also crucial for issues of architecture of, and connectivity within, the language system. Perhaps the most important of these issues is the balance between storage and computation: compositionality can often
 565 be rescued by increasing the demand on the storage component of the architecture, whereas it must be abandoned if one puts more realistic constraints on storage. In the latter case, of course, the demand on the computational component is increased.

Acknowledgments. We wish to thank Wilfrid Hodges, Karl Magnus Petersson and two anonymous reviewers for comments on earlier versions of this chapter. We are grateful to the Netherlands Organization for Scientific Research for support under grant 051.04.040.

REFERENCES

- Baggio, G. & M. van Lambalgen (2007). The processing consequences of the imperfective paradox. *Journal of Semantics* 24, 307–330.
- Baggio, G., M. van Lambalgen, & P. Hagoort (2008a). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language* 59, 36–53.
- Baggio, G., M. van Lambalgen, & P. Hagoort (2008b). Language, linguistics and cognition. In M. Stokhof & J. Groenendijk (Eds.), *Handbook of Philosophy of Linguistics*. Amsterdam-New York: Elsevier. In press.
- Barton, S. & A. Sanford (1993). A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Memory and Cognition* 21, 477–487.
- Borg, E. (2004). *Minimal Semantics*. Oxford University Press.
- Brown, C. & P. Hagoort (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience* 5, 34–44.
- Chauncey, K., A. Özyürek, P. Hagoort, & S. Kita (2004). Recognition of iconic gestures: A first gating study. Unpublished manuscript.
- Christianson, K., A. Hollingworth, J. Halliwell, & F. Ferreira (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42, 368–407.
- Croft, W. (1998). The structure of events and the structure of language. In M. Tomasello (Ed.), *The new psychology of language: cognitive and functional approaches to language structure*, pp. 67–92. Mahwah, NJ: Lawrence Erlbaum Associates.
- Culicover, P. & R. Jackendoff (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Culicover, P. & R. Jackendoff (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences* 10(9), 413–418.
- Dever, J. (1999). Compositionality as methodology. *Linguistics and Philosophy* 22, 311–326.

- Erickson, T. & M. Matteson (1981). From words to meanings: A semantic illusion. *Journal of Verbal Learning and Verbal Behaviour* 20, 540–552.
- Ferreira, F., V. Ferraro, & K. Bailey (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11, 11–15.
- Ferreira, F. & N. Patson (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass* 1, 71–83.
- Ferretti, T., M. Kutas, & K. McRae (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33, 182–196.
- Fodor, J. (1983). *The Modularity of Mind*. MIT Press.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance*, Volume XII, pp. 559–586. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frege, G. (1884). *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Koebner.
- Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78–84.
- Groenendijk, J. & M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.
- Hagoort, P. (2003). How the brain solves the binding problem for language: A neuro-computational model of syntactic processing. *Neuroimage* 20, S18–S29.
- Hagoort, P., G. Baggio, & R. Willems (2009). Semantic unification. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. MIT Press.
- Hagoort, P. & C. Brown (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing*, pp. 45–81. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hagoort, P., C. Brown, & J. Groothusen (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8, 439–483.
- Hagoort, P., C. Brown, & L. Osterhout (2001). The neurocognition of syntactic processing. In *The Cognitive Neuroscience of Language*, pp. 273–307. Oxford University Press.
- Hagoort, P., L. Hald, M. Bastiaansen, & K. Petersson (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441.
- Hintikka, J. (1983). *The Game of Language. Studies in Game-Theoretical Semantics and its Applications*. Synthese Language Library. Reidel.
- Hodges, W. (2006). From sentence meanings to full semantics. Unpublished manuscript.
- Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press.

- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Janssen, T. (1986). *Foundations and Applications of Mountague Grammar, Part 1*. Amsterdam: Center for Mathematics and Computer Science.
- Janssen, T. (1997). Compositionality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of Logic and Language*. Elsevier Science.
- Johnson, K. (2004). On the systematicity of language and thought. *Journal of Philosophy* 101, 111–139.
- Katz, J. & J. Fodor (1963). The structure of a semantic theory. *Language* 39(2), 170–210.
- Kazmi, A. & F. Pelletier (1998). Is compositionality formally vacuous? *Linguistics and Philosophy* 21, 629–633.
- Keenan, E. (1979). On surface form and logical form. *Studies in the Linguistic Sciences* 8, 163–203.
- Kim, A. & L. Osterhout (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language* 52, 205–225.
- Kutas, M. & S. Hillyard (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Kutas, M. & S. Hillyard (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Li, X., P. Hagoort, & Y. Yang (2008). Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in chinese. *Journal of Cognitive Neuroscience* 20, 906–915.
- Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- McElree, B., L. Pyllkanen, M. Pickering, & M. Traxler (2006). A time course analysis of enriched composition. *Psychonomic Bulletin and Review* 13(1), 53–59.
- McElree, B., M. Traxler, M. Pickering, R. Seely, & R. Jackendoff (2001). Reading time evidence for enriched composition. *Cognition* 78(0010-0277 (Print)), B17–25.
- Miller, G. & P. Johnson-Laird (1976). *Language and Perception*. Harvard University Press.
- Müntz, T., K. Schiltz, & M. Kutas (1998). When temporal terms belie conceptual order. *Nature* 395, 71–3.
- Nieuwland, M. & J. van Berkum (2005). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18, 1098–1111.

- Osterhout, L. & P. Holcomb (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31, 785–806.
- Özyürek, A., R. Willems, S. Kita, & P. Hagoort (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19, 605–616.
- Partee, B., A. ter Meulen, & R. Wall (1990). *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.
- Piñango, M., A. Winnick, R. Ullah, & E. Zurif (2006). The time course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research* 35, 233–244.
- Piñango, M., E. Zurif, & R. Jackendoff (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research* 28, 395–414.
- Pullum, G. & B. Scholz (2007). Systematicity and natural language syntax. *Croatian Journal of Philosophy* 7, 375–402.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pylkkänen, L., R. Llinas, & B. McElree (2007). An MEG study of silent meaning. Submitted.
- Robbins, P. (2007). Minimalism and modularity. In G. Preyer & G. Peter (Eds.), *Context-sensitivity and semantic minimalism*. Oxford University Press.
- Sanford, A. (2002). Context, attention and depth of processing during interpretation. *Mind and Language* 17, 188–206.
- Sanford, A. & P. Sturt (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences* 6, 382–386.
- Szabó, Z. (2007). Compositionality. Stanford Encyclopedia of Philosophy.
- Traxler, M., B. McElree, R. Williams, & M. Pickering (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language* 53, 1–25.
- Traxler, M., M. Pickering, & B. McElree (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language* 47, 530–547.
- Van Berkum, J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond*, pp. 229–270. New York: Psychology Press.
- van Berkum, J., P. Hagoort, & C. Brown (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience* 11, 657–671.

- van Berkum, J., P. Zwitserlood, P. Hagoort, & C. Brown (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research* 17, 701–718.
- van Lambalgen, M. & F. Hamm (2004). *The Proper Treatment of Events*. Oxford: Blackwell.
- van Oostendorp, H. & S. de Mul (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica* 74, 35–46.
- Westerståhl, D. (1998). On the mathematical proofs of the vacuity of compositionality. *Linguistics and Philosophy* 21, 635–643.
- Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy* 17, 329–342.