COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12949

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12949

Compositionality in a Parallel Architecture for Language Processing

Giosuè Baggio

Language Acquisition and Language Processing Lab, Department of Language and Literature, Faculty of Humanities, Norwegian University of Science and Technology

Received 23 March 2020; received in revised form 20 November 2020; accepted 27 January 2021

Abstract

Compositionality has been a central concept in linguistics and philosophy for decades, and it is increasingly prominent in many other areas of cognitive science. Its status, however, remains contentious. Here, I reassess the nature and scope of the principle of compositionality (Partee, 1995) from the perspective of psycholinguistics and cognitive neuroscience. First, I review classic arguments for compositionality and conclude that they fail to establish compositionality as a property of human language. Next, I state a new competence argument, acknowledging the fact that any competent user of a language L can assign to most expressions in L at least one meaning which is a function *only* of the meanings of the expression's parts and of its syntactic structure. I then discuss selected results from cognitive neuroscience, indicating that the human brain possesses the processing capacities presupposed by the competence argument. Finally, I outline a language processing architecture consistent with the neuroscience results, where semantic representations may be generated by a syntax-driven stream *and* by an "asyntactic" processing stream, jointly or independently. Compositionality is viewed as a constraint on computation in the former stream only.

Keywords: Compositionality; Semantics; Syntax; Language processing; Parallel architecture

1. Why cognitive neuroscience?

This paper takes a critical look at the principle of semantic compositionality (PSC) from the viewpoint of cognitive neuroscience. As defined by Partee (1995), the PSC is the thesis

Correspondence should be sent to Giosuè Baggio, Department of Language and Literature, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway. E-mail: giosue.baggio@ntnu.no

that "the meaning of a whole is a function of the meanings of the parts and of the way they are syntactically combined." It has been repeatedly pointed out that the PSC, in order to do any work at all, requires one to define at least what are the relevant "parts" of the whole, their "meanings," and their "syntactic combination." In other words, the principle can only be made precise within a theory of language, including at least a theory of lexical semantics and a theory of syntax (Groenendijk & Stokhof, 2004; Partee, 1984). In contemporary formal linguistics, the question has almost never been whether the PSC is true or false of given natural languages. The question has rather been what is needed for the PSC to hold, more specifically what theories of lexical semantics and syntax are sufficient to derive compositionally all and only the complex meanings sanctioned by our intuitions. The latter are given as the "data," compositionality is posited as a constraint on the derivation, and theories of syntax and semantics are designed and adjusted in order to satisfy the constraint while explaining the "data." What one can learn from this approach is which *theories* of linguistic meaning are compositional and which ones are not.

In order to learn whether—or better, as we shall see, to what extent—*language and meaning* themselves are compositional, a different strategy is required. Considering the PSC from the perspective of cognitive neuroscience allows us to approach (and begin to address) that question. The first move is to acknowledge that a prominent goal of theory development in linguistics and psycholinguistics is to capture syntax and semantics in the brain in explicit algorithmic terms (van Rooij & Baggio, 2021). As theoretical and empirical research progress, we hope to come to a more accurate picture of how the brain learns and processes languages, what types of structures it computes, and—as a function of all that—to what extent the PSC holds. That is the approach pursued here. The aim is to find out whether and how the PSC is true, not of particular languages, but of language as cognitive function.

The second move, necessary to motivate and develop a cognitive neuroscience take on semantic compositionality, is to identify some pragmatic assumptions implicit in the formulation of the PSC given above. The first concerns the interpretation of the definite NP "the meaning of a whole," which suggests the NP has a unique referent: Each "whole"—each expression: a complex word, a phrase, a sentence, and perhaps a discourse, too—has *one and only one* meaning. As linguists know well, this is not a plausible assumption to make. Any expression can have several different meanings: Only some may be derivable compositionally. Dropping the uniqueness assumption leads to two alternative versions of the PSC: either all meanings (PSC-A), or at least one meaning (PSC-O), of a complex expression is a function of the meanings of the parts and of their syntactic mode of combination. PSC-A implies a strengthening of the PSC. Indeed, linguists often seek a compositional account of the fact that a phrase or sentence has multiple (related or unrelated) meanings: either because some parts have multiple senses (e.g., lexical polysemy) or because alternative logico-syntactic analyses are possible (e.g., scope ambiguities). Instead, PSC-O implies a substantial weakening of the PSC, but again this is a possible theoretical position. For example, Borg (2004, p. 21) remarks that "every sentence in natural language has a meaning which is exhausted by the meanings of its parts and their mode of composition." It is the task of a theory of language to explain how *these meanings* are computed. It is an additional and separate set of questions whether there exist other types of meanings, what they are in each case, and how to account for them. As will be seen in Section 3, cognitive neuroscience may have something to say about these issues, too.

The second assumption of the PSC is the implicature that the meaning of a whole is a function *only* of the meanings of the parts and the syntactic mode of composition. Depending on whether or not one accepts this implicature, one gets again different notions of compositionality. The PSC minus the implicature (PSC − I) yields a weaker version of the PSC, which seems, however, sufficient to explain certain classic "facts" about language,[1] including *productivity* (i.e., that a user of a language L can produce and understand new sentences in L, if she knows the meanings of the "parts" of the new sentences and the syntax of L) and *systematicity* (that a user who understands a sentence in L can also understand syntactically related sentences in L, e.g., "John loves Peter" and "Peter loves John"). Borg (2012) notes that such facts "only support the claim that semantic content is *in part* determined via formal operations over syntactic forms" (p. 6). The PSC − I is silent as to what else may determine semantic content, and it does not impose any restrictions on that. Borg (2012) observes that it is a different issue whether "there exists an *exclusively* syntactic and lexical route to sentence-level semantic content," which is compositionality *plus* the implicature that phrasal and sentential meanings are determined *only* by lexical meanings and syntax (PSC + I). In this case, too, as I shall argue in Section 4, cognitive neuroscience provides data that speak to these issues.

So, what does a cognitive neuroscience approach to compositionality have to offer? The answer is a take on the PSC which is complementary to the one put forward in linguistics. To appreciate this point, we may describe formal and empirical theories of language as machines with fixed and moving parts. In linguistic theory, the fixed parts are the meanings of complex expressions (given by intuitions, judgments, and other relevant "data") and the PSC itself (a constraint on theories), while the moving parts are theories of syntax and semantics. In neuroscience, instead, the fixed parts are the meanings of complex expressions and partial knowledge of the organization of lexical semantics and syntax in the brain, while the moving part is the PSC itself. In both cases, we may learn about the moving parts as a function of assumptions or evidence about the fixed parts. Linguistics can tell us which theories of syntax and lexical semantics afford compositional analyses of particular linguistic phenomena. Neuroscience can tell us to what extent specific versions of the PSC may hold, given the organization of relevant systems of syntax and semantics in the brain. Fleshing out the pragmatics of the PSC yields a partial list of versions (PSC-A, PSC-O, PSC + I, PSC – I, etc.) that we may then try to assess empirically, as we would with most other principles of human language processing (Baggio, van Lambalgen, & Hagoort, 2012).

2. Why compositionality?

Is the cognitive neuroscience approach outlined here far removed from other views of compositionality in the cognitive sciences? Is it at all legitimate to restate the PSC as a principle of human language processing? At first blush, this would seem like a misconception of what the PSC is supposed to mean and do, at least if viewed from the decidedly anti-psychologistic perspective on semantics in the tradition of Frege, Tarski, Montague, Lewis, and others (for a classic statement, see Lewis, 1972, p. 170). But a closer look at standard arguments for the PSC indicates that psycholinguistic considerations are often implicitly used in the reasoning. So let us take a step back and review some of those arguments.

There are two types of arguments for (or against) compositionality: (I) arguments that start from supposed "facts" about language, or meaning, thought, mind, etc., and that reach conclusions of the form "the PSC is (not) required to explain those facts"; (II) arguments that, instead, justify compositionality by stating the functions or roles it plays in certain theoretical or empirical endeavors, for example, in linguistics, philosophy, computer science, etc. There may well be overlap between some of these strategies, and arguments for or against the PSC may be designed that mix elements of (I) and (II). Below, however, I will consider some "pure" type-I and type-II arguments.

The standard type-I arguments for the PSC—for example, learnability, novelty, productivity, systematicity, communication, and so on—have all been found wanting as justifications of the *necessity* of compositionality. Pagin and Westerståhl's (2010b) pithy treatment of these arguments reaches the dismaying conclusion that they all only establish that the semantics—that is, the meaning composition operations that, in standard analyses, mirror syntactic structure building—has to be *computable*. However, computability only requires a recursive semantics, on top of a recursive syntax. But one can easily construct recursive, computable theories of meaning that are *not* compositional: for example, where the semantic operators do not track syntactic structure, or in which extralinguistic information is used to derive meanings in context. Importantly, even though such a recursive and computable semantic theory may not formally satisfy the PSC, it could still explain: How one can acquire a language where infinitely many meanings can be expressed ("learnability"); how one can understand ("novelty") and produce ("productivity") infinitely many sentences, including the infinite set of new sentences; how one, who understands expressions of the general form *aRb*, can also understand expressions of the form *bRa* ("systematicity"; see above); and why, often, the sender and receiver of linguistic messages converge on shared interpretations ("communication"). Computability does not entail compositionality. This point can be made through formal arguments (Pagin & Westerståhl, 2010a) or concrete examples (see recent work using artificial neural networks; Baroni, 2020; Nefdt, 2020). Then, arguments that establish that meaning is or should be computable do not necessarily also establish the PSC as a logical consequence of computability. Compositionality is a narrower, stronger formal notion than computability. As such, it requires specific arguments that will differ from standard ones, to the extent that those only establish the computability of meaning, and in spite of their ability to explain additional facts such as learnability, novelty, productivity, and so on (Baggio et al., 2012; Werning, 2005).

There is a type-I argument, offered by Pagin and Westerståhl (2010b), which appears more convincing. They call it the "complexity argument." Why should the semantics be compositional, in addition to being computable? One answer may be that formal computability is a broader and weaker concept than *tractability*, which is obviously required to ensure that the meanings of complex expressions "can be computed by ordinary language users during the limited time they usually have" (p. 269; this is the psycholinguistic consideration, crucial for the argument to go through) and, one might add, given the finite memory resources they have. Pagin and Westerståhl argue that "compositionality is not sufficient for low complexity." Still, a stronger version of the PSC, where semantic composition operators are *polynomial-time computable* might suffice. Unfortunately, polynomial-time computability (or even less restrictive tractability standards; see van Rooij, 2008) does *not* entail compositionality. We can conceive of operations which are polynomial-time computable (or fixed-parameter tractable), but do *not* track syntax, as the PSC requires. Tractability is desirable, but it is not sufficient to establish the PSC (see below). The "complexity argument" fails, too. This conclusion is compatible with the view that certain reasonable complexity restrictions on composition operators will entail tractability of the semantics. What is at issue here is rather the inverse inference, from tractability to compositionality: If Pagin and Westerståhl are right, the former does not entail the latter. The upshot is that type-I arguments fail (qua arguments for compositionality), because they only establish the computability or the tractability of composition. But as neither entails compositionality, these arguments do not warrant the PSC.

Let us consider the most significant type-II arguments. One prevalent argument for compositionality is the "methodological argument." It is a kind of pragmatic case for the PSC, and herein lie both its weakness and its strength. These arguments usually start from the premise that the PSC is formally vacuous: Any grammar can be given a compositional semantics (Janssen, 1986, 1997; Zadrozny, 1994).[2] This also entails that the PSC is empirically vacuous (Baggio et al., 2012): If one can always produce compositional analyses of given linguistic structures, then compositionality is never challenged by empirical (i.e., linguistic, psycholinguistic etc.) data. These arguments would point to a non-substantive notion of compositionality. Groenendijk and Stokhof (2004) then argue that the PSC "must be viewed as a methodological principle, one that represents a choice to do semantics in a particular way." This view is standard in many corners of semantics, but it may be framed in more or less radical versions. Provocatively, Janssen (1997) writes that "If a proposal is not compositional, it is an indication that the fundamental question what the basic units are, is not answered satisfactorily." Then he adds: "So the main reason to follow this methodology, is that compositionality guides research in the right direction." A more moderate account of the PSC as a methodological principle is that it may serve as "an arbiter between rival analyses of a given set of phenomena" (Groenendijk & Stokhof, 2004): A classic case is the development of dynamic predicate logic as a compositional alternative to (non-compositional) discourse representation theory (see also Szabó, 2000). A weaker view still is that the PSC does not actually constrain semantic theory much, but it "makes for elegance and uniformity of presentation" (van Benthem, 1986). It is clear to anyone working in formal semantics or carefully watching its operations that the "standard view" of the PSC is, indeed, that of a "methodological standard by which formal semanticists go about their business" (Groenendijk & Stokhof, 2004).

The "standard view" goes back at least to Tarski (1956), who called it the "recursive method." As reconstructed by Hintikka and Sandu (1999), this amounts to "defining a semantical concept[3] first for elementary expressions, and then extending it step by step to complex ones in tandem with the syntactical operations that generate these complex expressions." This approach, which presupposes the PSC, "is truly Tarski's 'methodological secret' in his work on truth-definitions" (Hintikka & Sandu, 1999). Tarski believed that one could provide definitions of semantic concepts (e.g., truth) for "formalized languages," such as logics, but not for natural languages. It is only in Montague's work that an extension of the "recursive method" and, with it, of the PSC to natural languages is granted, on the basic assumption that natural languages are not unlike formal languages ("I reject the contention that an important theoretical difference exists between formal and natural languages"; Montague, 1970, p. 189). But then, just *what is compositional* according to the methodological argument? It is not natural languages as such, but rather the logical object languages into which we translate natural language sentences and, therefore, the theories resulting from that. The methodological argument is a forceful pragmatic justification of the PSC for the working semanticist. But it is also a stark reminder of what one may expect to learn from formal semantics: not whether natural languages are (not) compositional, but whether particular semantic theories accord with particular versions of the PSC.

But suppose one really wants to learn whether (a fragment of) a natural language is compositional. In that case, the methodological strategy would just not do. All it can do is to supply us with techniques to secure compositionality,[4] even if there is none "in" the given natural language. As stated with great clarity by Szabó (2000), on the "standard view," "when we say that a certain theory of English is compositional, we mean that the theory presents English as a compositional language." But is English *itself* compositional? Now we are effectively asking whether the PSC would count as an empirical hypothesis about a fragment of a natural language (e.g., adjective–noun phrases in English), a language (English as a whole), all (possible) human languages (Szabó, 2000), programming or other artificial languages, animal signaling systems, and so on. This is the idea of empirical type-II arguments for the PSC.

There are at least two ways of interpreting the PSC as an empirical hypothesis. One views questions of the form "Does L [a language or fragment] accord with the PSC?" as allowing yes/no answers. To avoid making "yes" answers exceedingly more likely than "no" answers, one has to find or devise "independently motivated constraints" (Groenendijk & Stokhof, 2004) on theories of syntax and (lexical) semantics. If such theories are allowed to vary freely, then one is free to make them vary so as to fulfill the PSC. But then we are back to the methodological strategy, and any "yes" answer may just mean that the theorist has been clever enough to produce a compositional treatment of L, but not that L is compositional. This is not the empirical procedure of discovery that we are after. The second interpretation is suggested by Dowty (2007): "Compositionality really should be considered "an empirical question." But it is not a yes-no question; rather, it is a "how"-question." Dowty goes on to give a definition of "natural language compositionality" (NLC):

> Whatever strategies and principles we discover that natural languages actually do employ to derive the meanings of sentences, on the basis of whatever aspects of syntax and whatever additional information (if any) research shows that they do in fact depend on. [NLC]

This is obviously a major departure from the PSC, in both its letter and spirit. First, NLC seems to contradict the PSC + I (the standard PSC reading), because it allows for any "additional information," besides syntax, to contribute to sentence meaning. But it is also not clear how *empirical* NLC is, that is, whether it is at all possible for it to fail. Dowty (2007) writes: "Under this revised terminology, there can be no such things as 'counterexamples to compositionality', but there will surely be counterexamples to many particular hypotheses we contemplate as to the form that it takes." That is, the NLC is not going to be found true or false ("it is not a yes-no question"). It is the particular analyses proposed in each case that may be correct or incorrect accounts of the "strategies and principles" by which sentential meanings are actually derived ("it is a "how"-question"). Dowty is quick to preempt an uncharitable "anything goes" reading of his proposal and appeals to considerations of "transparency," "simplicity," and "economy" as criteria to evaluate hypotheses about NLC. Given the "prevalence of straightforwardly compositional linguistic data," one may assume that languages are, generally, or as a default, straightforwardly compositional. The issue is "where, exactly, transparent compositionality *stops* (if it does) and how compositionality [it is NLC, not the PSC, that Dowty has in mind here] works from there on."

One reading of Dowty (2007) is that there exist "lower-complexity" phenomena that we may capture via fairly conservative compositional analyses, where syntactic and semantic composition are "no more complicated than they need to be" and there is a transparent syntax–semantics correspondence (Dowty, 2007). But there are other, "higher-complexity" phenomena that we may not be able to study in the same ways. The PSC—specifically, the PSC + I—may cover the first set of phenomena, while NLC extends to the uncharted territory of the second set. For Dowty (2007), *all of this is compositional*: The quotation above is revealing ("compositionality stops (if it does) and (...) compositionality works from there on"). However, Dowty's vague modifier "straightforwardly compositional" is problematic, given the well-known suppleness of the PSC. His argument just does not seem to suggest the kind of boundary on the PSC that would make it a bona fide empirical hypothesis.

Now let us suppose, for the sake of argument, that Dowty's view actually does set a boundary on the PSC ("transparent compositionality"). It should be said that such a boundary would be *internal* to the class of algorithms computing (polynomial-time, or in any case tractable) functions of simpler-to-complex meanings. Language users should be able to compute the "meaning of a whole" in limited time and using finite memory resources, even for non-straightforwardly compositional phenomena. The higher-complexity phenomena of Dowty (where analyses are not fully transparent, simple, and economical) are still "low complexity" phenomena in the sense of Pagin and Westerståhl (2010b): By definition, they all fall in the class of tractable problems. But as said for Pagin and

Westerståhl's (2010b) "complexity argument," *computability does not entail compositionality*. Dowty's case, therefore, collapses into an argument for the computability of meaning, as is apparent from his formulation of NLC, plus a problematic "fuzzy" line separating lower-complexity from higher-complexity, but in all cases finitely computable, composition problems.

One more type-II argument, before we turn to some ideas from neuroscience. There is a possible argument for a version of the PSC, which one may call the "competence argument." The argument treats the PSC as being applicable to language as cognitive function, or to linguistic competence, to the extent that the PSC constrains semantic processing in the brain. A competent user of a language L is able to compute *at least one* meaning or interpretation $M(e)$ for some expressions e in L, such that $M(e)$ is a function *only* of e's syntactic structure and of the meanings of e's parts (see Pagin & Westerståhl, 2010a, 3.1–3.2 for explicit definitions). Compositionality is an abstract, high-level, general account of precisely this aspect of linguistic competence. This is a mix of the PSC-O, which makes the formulation relatively liberal, and of the PSC + I, which rebalances it along more conventional lines (see also Borg, 2004, 2012). The "competence version" of the principle of semantic compositionality (PSC-C) is:

> At least one meaning of a whole is a function only of the meanings of the parts and of the way they are syntactically combined. [PSC-C]

PSC-C requires us to show that any competent user of a language L is able to assign, to at least *some* (not all) expressions e in L, different meanings, and that at least one of those meanings is as specified by the PSC + I. Moreover, the new principle, as any other version of the PSC that strives to be a cognitively plausible hypothesis of how humans compute meaning, must be bound by further constraints on the tractability of syntactic and semantic composition operators, which should be compatible with finite computational resources in the brain (van Rooij & Baggio, 2021).

Linguistic phenomena which have long been considered as counterexamples to the PSC include idioms (1), metonymy (reference transfers) (2), contextually restricted quantification (3), constructions that pick out the endpoint on a relevant scale, such as superlatives and ordinals (4), and others:

1. John spilled the beans just before dinner.
2. The wax museum is displaying The Beatles.
3. When it snows in Bergen, everyone is happy.
4. The first case of AIDS was reported in 1975.

Presumably, competent users of English *can* assign compositional meanings to each of (1)–(4), as required by the PSC-C: For example, one *could* interpret (2) as being about Paul and Ringo's guest appearance at a wax museum, and (4) as saying that the first case of AIDS to ever have occurred was reported in 1975. These compositional meanings may or may not be contextually appropriate for *utterances* of (1)–(4). But this just shows that the PSC-C is not a comprehensive account of linguistic competence. The PSC-C is necessary but not sufficient. A complete theory should explain not just how language users can assign alternative (non)compositional meanings to expressions. It should also state in what conditions compositional interpretation is engaged, and it should explain how one can produce and compare different interpretations of the same expressions. I will develop further this line of argument in Sections 3 and 4.

3. Compositionality and representational coexistence

The competence version of the PSC (PSC-C) requires us to show that some complex natural language expressions have multiple meanings, and that at least one of those meanings is a function *only* of the meanings of the parts and syntax (i.e., PSC-O plus PSC + I). If that is what a competent language user can do—assign compositional and non-compositional meanings to the same expressions in a language—then *how* do they do that? PSC-C has two empirical implications: (H1) that the brain can support compositional *and* non-compositional representations of the same expressions, and (H2) that there is "an *exclusively* syntactic and lexical route" (Borg, 2012) to phrasal and sentential meaning—possibly one of several such "routes to meaning." I address the first implication in this section and the second in Section 4.

I call "representational coexistence" (ReC) the human brain's (hypothetical) capacity to represent the same entities[5] compositionally *and* non-compositionally, that is, as a function of distinct representations of the parts and of their configuration, *and* by means of some alternative representational format. To assess whether evidence for ReC in humans also provides support for PSC-C, one should address two questions: First, whether evidence for ReC in a given domain of perceptual or cognitive entities (e.g., visual objects, memories of events, etc.; see below) could also apply to the case of language; second, whether non-compositional representations in each domain of entities are holistic or are instead a function of representations of the object's parts, of their configuration, *and some other content* (PSC − I). Importantly, PSC-C applies to language primarily. So, as far as the first question goes, we are seeking evidence for ReC for linguistic or language-like structures, or for other cognitive representations that linguistic meanings are plausibly based on. As for the second question, PSC-C is neutral with respect to the non-compositional meanings that may be computed by the system. In natural languages, we find both holistic meanings and meanings that are not strictly a function of word meanings and syntax; see examples (1)–(4). Whether evidence for ReC supports PSC-C does not hinge on the types of non-compositional meanings that the system is capable of generating, so long as the compositional ones are strictly a function of constituent meanings and syntax (PSC + I).

Is there evidence for ReC from human neuroscience? It has been argued that *visual processing*, for example object recognition, is "compositional" (Battaglia, Borensztajn, & Bod, 2012). The ventral visual system is *hierarchically organized*: Areas higher up in the cortical hierarchy (i.e., further away from V1) present more complex responses by virtue of computing combinations of features coded at lower levels of the hierarchy. It is not clear how this "combinatoriality" relates to compositionality, in the sense of the PSC. Compositionality indeed affords "new" combinations of "old" parts (even if the PSC is not necessary to explain this fact; see above), but that is precisely the challenge for convergent, feedforward, hierarchical network architectures. The challenge here is the "combinatorial explosion" of the total number of higher-order neurons required to code all possible, or all discriminable, combinations of features, as they appear in different visual objects or in different presentations of an object across modalities. Higher-order neurons may fire *selectively* in response to particular combinations of lower-level features, which limits the number of different combinations of features that each cell may respond to. Some types of visual objects may be recognized using this mechanism: faces, for example (Perrett, Rolls, & Caan, 1982; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994). Different binding mechanisms, however, are needed to represent new feature combinations.

Population coding may provide the requisite mechanism. The same neurons may be active in a given process within one cell assembly and in a different process within another cell assembly. Neurons coding for feature X could form a cell assembly with cells coding for feature Y, thus representing the feature combination [X Y], and they may later join in another assembly with neurons coding for feature Z, representing the combination [X Z]. This solution would work only if neurons could be "labeled" dynamically as belonging to the same cell assembly, and only if those "labels" could be switched "on" and "off" rapidly. Synchrony of firing and enhanced firing rates have been indicated as possible mechanisms for "labeling" units in neuronal populations. This proposal solves the problem of representing new objects: In principle, any two sets of neurons, coding for basic features, could be synchronized so as to produce a new percept. But population coding may not work as well, if the system is trying to represent multiple objects sharing some features (e.g., [[X Y] [X Z]], red square and red circle), and if multiple cell assemblies must be formed simultaneously, engaging some of the same cells. This "superposition problem" could be solved by segregating cell assemblies in time, or by generating different "tokens" of the same features.

Both hierarchical feedforward architectures and population coding models of visual binding can indeed support *structured representations* of visual objects. However, it is not clear that the limited combinatorial capacity of the primate's visual system is sufficient for compositionality, in the sense of the PSC-C. Combinatoriality in vision does not lend support to (H1)–(H2) and cannot satisfy the two requirements above: Structured representations of visual objects are *not* language-like (e.g., their spatial configuration cannot count as syntactic), and most importantly there is no evidence for the coexistence of "compositional" (e.g., featured based) and "non-compositional" (e.g., holistic) representations of the same visual objects. Research in vision science only provides suggestive evidence for the existence of multiple neural mechanisms for the combination of visual features into structured representations.

Consider now another cognitive domain where representations have been regarded as "compositional": *declarative memory*. James (1890) noted that events in memory could be either *fused* (differences between individual constituent items are lost) or *associated* (individual items are differentiated and bound together). Evidence from animal and human research shows that memories of events may be encoded in two "formats" (Cohen & Eichenbaum, 1993; Eichenbaum, Otto, & Cohen, 1994; Eichenbaum, Schoenbaum, Young, & Bunsey, 1996; Hannula, Ryan, & Warren, 2017; Henke, 2010; Rubin & Cohen, 2017; Rubin, Schwarb, Lucas, Dulas, & Cohen, 2017):

a. A *configural format*, where the event elements are fused, or compressed into a holistic, part-free representation; the event elements, if presented in isolation and in new contexts, will not be generally recognized as part of that event; such configural representations may play a functional role in the acquisition of biases or adaptations to particular events;
b. A *compositional format*, where the event elements are encoded together with relevant relationships (spatial, temporal, etc.) between them; event elements can be recognized individually or relationally when presented in isolation and in new contexts; thus, event memories can be expressed (retrieved, modified, etc.) in novel situations.

In humans and in other mammal species, representations in these two "formats" are generated and maintained by distinct brain circuits and mechanisms. *Compositional representations*, which may eventually be stored as declarative memories, originate in interactions between the hippocampus and prefrontal cortex. The hippocampus binds together elements of experience that may co-occur arbitrarily and only once, while prefrontal regions contribute to maintenance of representations of elements of experience, to gradual abstraction from contextual information, and to connecting together memories of the same type. In contrast, *configural representations*, which may result in implicit, non-declarative memories, originate in the parahippocampal cortex and in the neocortex. The joint activity of such structures allows a degree of associative processing, but here binding results in fused, inflexible representations of complex experiences. Therefore, the hippocampus is crucial for the organization of declarative memories in terms of flexible relations between parts and wholes.

We reach the same conclusion we arrived at for vision: There exist several routes or neural mechanisms for producing representations of the same type, that is, memories. But, in this case, there is more: The hippocampal system-dependent route produces compositional representations; the hippocampal system-independent route instead yields non-compositional representations. This may seem relevant to ReC and (H1), but it is not quite it: (H1) entails that the brain can support both compositional and non-compositional representations *of the same entities* (of the same objects, events, etc.), and not just of the same *types* of representations (of objects, events etc.). (H1) requires, for a memory content C (e.g., of an episode, or event), that we encode and store a compositional *and* a configural memories with content C. There is evidence for redundant representations of items in working memory tasks (for a review, see Tamber-Rosenau & Marois, 2016; see also below). But no such evidence is available for declarative memory. In fact, the cortico-hippocampal memory system embodies mechanisms that preclude encoding of redundant information and routinely erase redundant memories (Frankland & Bontempi, 2005). Further, the compositionality of C is one *condition* on one's having a (declarative) memory with content C (Cheng & Werning, 2016). If C is not compositional, the result is not a declarative memory. Each event may be encoded *either* as a declarative memory (compositionally) *or* as a non-declarative memory (configurally). If this is correct, the exclusive disjunction is inconsistent with (H1) and ReC.

There is support for (H1) and ReC from a functional magnetic resonance imaging (fMRI) study by Reverberi, Görgen, and Haynes (2012a) on the neural representation of logical decision rules. They used simple rules:

S1: If there is a house press left.
S2: If there is a face press right.
S3: If there is a face press left.
S4: If there is a house press right.

and compounds of simple rules: C1: S1 + S2; C2: S3 + S4. In the scanner, in each trial, participants had to recall and maintain one S- or C-rule. After a variable delay, they had to respond to a visual target (e.g., the image of a house) by pressing the correct key, according to the rule. Using multivariate pattern analysis on activation vectors, they found two cortical regions that contained information on C-rules: right frontal and left parietal cortex. In their "compositionality analysis," they trained classifiers on vectors from S-rules, and they tested them on vectors from C-rules. In the right frontal area, but not in the left parietal area, C-rules could be decoded from S-rules. In their "inverse compositionality analysis," they trained classifiers on vectors from C-rules, and they tested them on vectors from S-rules. Again, in the right frontal cortex, but not in the left parietal cortex, S-rules could be decoded from C-rules. The two areas seem to "use a different neural code to represent the same information" (Reverberi et al., 2012a, p. 1242): a "compositional code" in the right frontal cortex; and a "unique," "independent," non-compositional code in the left parietal cortex.[6]

These data lend credence to the notion of ReC, confirming implication (H1): The brain can represent compositionally and non-compositionally *the same entities*. This differs from the case of the memory system: There, events can be encoded *either* compositionally *or* configurally; instead, logical decision rules can be encoded compositionally *and* non-compositionally in different cortical areas. This is attested in humans, but not in monkeys. So far, there is evidence for "independent coding" of items in the monkey prefrontal cortex, but not for "compositional coding" (Warden & Miller, 2007, 2010). Studies by the same group (Baggio et al., 2016; Reverberi, Görgen, & Haynes, 2012b) provide further evidence that a bilateral network of prefrontal and parietal areas supports multiple coexisting representations of the same logical decision rules, as a function of task-relevant structural or semantic aspects of those rules: One finds evidence for distinct cortical regions representing the surface form versus the logical meaning of rules, and rule identity versus rule order.

The notions of "compositionality" encountered in this section do not entirely capture the PSC. In studies of vision, memory, and decision rules, "compositionality" entails that representations of parts are *recoverable* from representations of wholes,[7] that some representations of wholes are functions of representations of their parts. But the PSC stipulates: *and of the way they are syntactically combined*. The PSC is meant to apply to linguistic (semantic) competence: ReC and (H1) must then be assessed in the context of language processing research, too. Idioms are a classic test case.

Debates about idiom comprehension have revolved around the questions whether, when, and how compositional (or literal) meaning is used during idiom processing (Vulchanova, Milburn, Vulchanov, & Baggio, 2019). It is generally accepted that the literal meanings of some expressions, occurring in idiomatic phrases (e.g., “beans” in “spill the beans”), “linger” during the retrieval or construction of idiomatic meaning. This would suggest that the compositional meaning of an idiomatic phrase may be partly available, too. The question then is whether such (partial) compositional and idiomatic meanings are computed sequentially or in parallel, and, if sequentially, in what order. Evidence of parallel activation would count as evidence for ReC.

The configuration hypothesis of idiom processing is one prominent and influential sequential model (Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988). The idea is that interpretation may proceed largely compositionally until the comprehender recognizes that the phrase they are processing is an idiom, that is, until the “idiom key” is reached. Then, the figurative meaning is activated, and compositional processing halts. There is supporting evidence for this account, but also data points that conflict with it: Smolka, Rabanus, and Rösler (2007) show evidence for the activation of the literal meanings of sentence-final verbs in idioms, after the phrase has been recognized as an idiom. In contrast, several parallel models posit two routes or mechanisms subserving the activation and the construction of compositional and figurative meaning (Carston, 2010; Swinney & Cutler, 1979; Titone & Connine, 1999). Under the hybrid model by Titone and Connine (1999), idiom comprehension follows two simultaneous parallel streams: direct access of figurative meaning, as soon as the idiom is identified; and meaning composition based on the literal meanings of the idiom’s constituents and syntax (PSC). These two streams may interact continuously, giving rise to effects of facilitation and interference in the construction of the intended idiomatic meaning. Titone and Connine (1999) demonstrated that compositional meaning facilitates the activation of idiomatic meaning, if the two are related, as in decomposable idioms.[8] Instead, compositional meaning may interfere with the activation of idiomatic meaning, if the two are unrelated, as in non-decomposable idioms. These effects are reflected in faster and slower processing of decomposable and non-decomposable idioms, respectively. Further evidence has come from production studies (Nordmann, Cleland, & Bull, 2013; Sprenger, Levelt, & Kempen, 2006). These results suggest that compositional and non-compositional meanings coexist on-line in brain time, not just in brain space, as the fMRI experiments on decision rules indicate.

Together, the selection of studies reviewed here indicates the following preliminary conclusions (Table 1). Evidence of multiple routes or mechanisms (MM) subserving the combination of parts into wholes is provided by research on object vision and declarative memory. Evidence for the existence of neural systems that can support both compositional and non-compositional representations (C/NC) has come from studies of declarative memory, whereas evidence that such representations may be constructed *for the same*

Table 1
Summary of processing and representational capacities of neural systems underlying four cognitive and perceptual functions. MM: multiple mechanisms/routes to structured representations; C/NC: support for compositional versus non-compositional representations of different entities; C/CN-S: support for compositional and non-compositional representations of the same entities; MM-C/CN-S: multiple mechanisms/routes to compositional *and* non-compositional representations of the same entities. Dots represent attested capacities; empty cells denote absence of positive evidence

Function	MM	C/NC	C/CN-S	MM-C/CN-S
Object vision	•			
Declarative memory	•	•		
Logical decision rules		•	•	
Natural language semantics			•	•

entities (C/CN-S) is provided by fMRI research on decision rules. Finally, psycholinguistic studies of idiom processing have produced evidence for the existence of two parallel and interacting processing streams, subserving the on-line computation of compositional and non-compositional meanings. Overall, we find some empirical support for ReC, for (H1)–(H2), and for the competence version of compositionality (PSC-C). These data challenge versions of compositionality that are not encompassed by PSC-C, in particular the combination of PSC-A and PSC + I. A formal theory where all meanings in language are generated strictly as a function of constituent meanings and syntax, and nothing else, cannot explain the existence of holistic meanings (e.g., of non-decomposable idioms) and meanings that depend on information other than the semantic values of constituents and syntactic structure. Such a theory would also be unable to connect to basic facts about the architecture of language processing, as revealed by studies of idioms. The choice is between two types of explanatory theories, based on either PSC-C (which accepts the distinction between, and coexistence of, compositional and non-compositional meanings) or a version of Dowty's NLC (where all meanings are compositional) that addresses the problems pointed out in Section 1. In the next section, the former option is pursued, in the context of a parallel architecture for language processing.

4. Compositionality in a parallel processing architecture

The preliminary conclusion that we reach, based on the discussion so far, is that the human brain can represent entities (e.g., memories of episodes, logical rules, natural language expressions) both compositionally and non-compositionally. In some domains, such as language and semantics, there are multiple parallel routes or neural mechanisms for constructing such representations. I will now discuss the role of compositionality (PSC-C) in a cognitive architecture that embodies such more general representational and processing capacities of the human brain.

In the neurolinguistics community, there is convergence on the view that linguistic information—phonological, orthographic, syntactic, semantic, and so on—is processed by multiple "streams" which can, to an extent, operate in parallel (for recent discussions, see Baggio, 2018; Friederici, 2017; Hickok & Poeppel, 2016). A parallel architecture for language processing has been developed at the intersection of generative syntax and cognitive linguistics (Jackendoff, 2007). Further, joint learning and joint parsing models have been used in computational linguistics (Surdeanu, Johansson, Meyers, Màrquez, & Nivre, 2008). In many of these models, syntactic and semantic representations are derived simultaneously in parallel, in contrast with traditional models in which a syntactic parse is typically computed first, and then fed to an interpretation or inference engine. The two-step, serial idea, encapsulated by the axiom "syntax proposes, semantics disposes" (Cram & Steedman, 1985, p. 325), is no longer considered a viable architectural hypothesis in many areas of the language sciences. Interestingly, while compositionality is often assumed to require such a serial architecture—such that constituent meanings and syntactic structures are fed into the semantic composition process—the traditional view of the PSC as syntax–semantic homomorphism is consistent with some degree of autonomy between syntax and semantics, considered as independent "modules."[9] Thus, there is room for the PSC even in a parallel architecture for language.

Below, I present a parallel architecture for language processing that is plausible and supported by experimental results, and that moreover assigns a definite role to the PSC-C (for a review of the evidence, see Baggio, 2018; for a computational analysis, see Michalon & Baggio, 2019; for other, partly compatible proposals, see Jackendoff, 2007; Kuperberg, 2007; Bornkessel-Schlesewsky & Schlesewsky, 2008).

The model posits two parallel streams for meaning and grammar, or "M-stream" and "G-stream," for short. The diagram in Fig. 1a shows the possible processing routes for lexical information (**L**), activated upon presentation of the input (*i*), for example, a single word in the context of an utterance. The lexicon includes phonological, orthographic, morphological, syntactic, semantic, and pragmatic features, associated with a visual or an auditory word. Depending on specific factors (see below), lexical information may be processed primarily in the M-stream, or in the G-stream, or in both streams in parallel. In all cases, the goal of computation is to update the current sentence or discourse model (**M**) with the information provided by the input. Processing occurs in two phases in both streams: Representations of contextual semantic relations (**R**) and syntactic structure (**S**) are updated first, before the model (**M**) itself is updated.

Lexical information is processed in parallel and simultaneously in the two streams. However, the *balance* between the two streams is not fixed and even, but evolves as novel inputs come in. So, at every new word in a sentence—but the same applies to smaller units, such as morphemes, or larger units, such as phrases—the balance can shift between three types of states: favoring the M-stream, favoring the G-stream, or balancing evenly the two. Only in particular circumstances does the balance remain fixed throughout a sentence, or is one of the two streams effectively "shut down": For example, Jabberwocky-style "sentences" may be parsed (G-stream), even though they convey no conceptual or referential meaning (M-stream). In normal circumstances, the balance between the two streams evolves in time, without one stream prevailing over the other, and without either stream being ever shut off in the process.

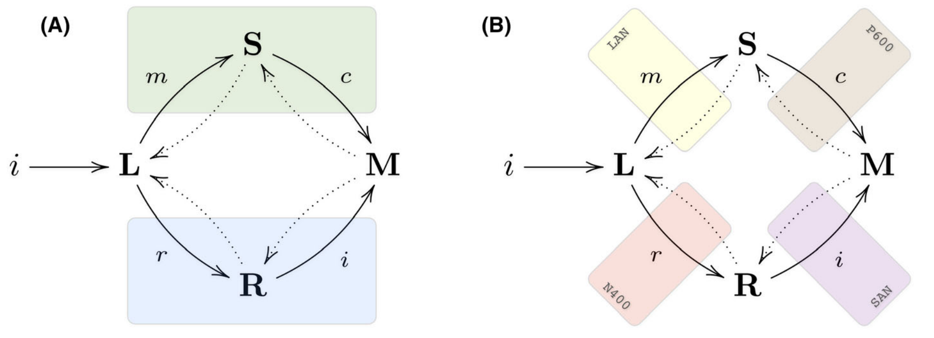

Fig. 1. (a) Diagram representing the processing of an input *i* (e.g., a word) in a cognitive architecture of the language system with parallel streams for meaning (blue) and grammar (green): L is lexical information activated by *i*; S is a structural (syntactic) representation of *i* in context; R is a relational semantic representation of *i* in context; M is the sentence (or discourse) model in which *i* is integrated; *m* is morphosyntactic processing; *c* is syntax-driven composition; *r* is relational semantic processing; *i* is interpretive processing. (b) At each input processing stage, bottom-up (thick arrows) or top-down (dotted arrows) processes may be reflected in modulations of different components of the event-related potential (ERP) signal: the N400 (orange shaded area) and the SAN (or Nref; purple) in the meaning stream; the LAN (yellow) and the P600 (brown) in the grammar stream.

One key factor, which determines the balance between streams, is which subset of the lexical features associated with the input word carries more "weight" for the purposes of updating the discourse model. For content words, semantic features (conceptual, referential, pragmatic) are critical. For example, even subtle differences between a pair of near-synonym nouns may make a difference for comprehension. Moreover, different senses of a word should be selected in order for comprehension to occur. Consider complex nominals: "color" in "color television," "color palette," "color consultant"; "musical" in "musical clock," "musical comedy," "musical criticism"; "flat" in "flat tire," "flat note," "flat beer," and so on (Braisby, 1998; Keenan, 1978; Lewis, 1986; Partee, 1984). In these cases, relevant semantic features should be extracted from the lexical representation **L** and processed by the M-stream. The hypothesis here is that content words engage primarily the M-stream, while the balance may shift, and increasingly involve the G-stream, as a function of the interpretive weight that morphological and syntactic features, associated with input content words, carry at a particular stage of processing (more below). In the M-stream, processing comprises two phases: a first phase, during which lexical-semantic features of the input word are related (*r*) to the lexical-semantic features of other, already processed content words in the discourse; the result is a relational representation (**R**) of the input word in its semantic context, encoding logical, associative, categorial, or other semantic relations; a second phase, where referring expressions (nouns, pronouns) and referentially salient expressions (e.g., some adverbs or prepositions) are interpreted (*i*) via assignment of values to variables coding for numerosity, spatial, and temporal coordinates for the entities or events described; logical or pragmatic inference, and forms of semantic enrichment (e.g., figurative processing), may contribute to updates of **M**. M-stream processing is "asyntactic" (Jackendoff, 2007), in the sense that, in general, it is not constrained by rules and regularities of grammar, which are the remit of the G-stream. This entails that the PSC does *not* apply to the M-stream, as such. Processes and representations in the M-stream, considered before **M** is updated, are organized according to independent conceptual, referential, and pragmatic principles of interpretation.

Input words or morphemes are processed primarily by the G-stream, if information relevant to interpretation is encoded morphologically or syntactically in the input's surface form, as is the case for function words, grammatical (non-root) morphemes (e.g., tense and number), and any lexical cues to specific grammatical constructions. In the G-stream, too, processing of grammatical features encoded in **L** goes through two phases: a first phase, in which morphosyntactic constraints are checked (m) for satisfaction (e.g., word category constraints, morpheme ordering, agreement), given the preceding sentence fragment; the result is a structural (syntactic) representation of the input word in context (**S**); a second phase, where the syntactic representation of the word in context is used to update, via syntax-driven composition (c), the discourse model **M**. This composition process is constrained by the PSC. As a first approximation, it may be captured by one or several logico-syntactic operations on lexical representations of elementary expressions (e.g., as functional application, Heim & Kratzer, 1998, or equivalent), although the proper choice of syntactic representations and operations is largely an empirical issue (Section 1). Two remarks: (a) These representations are abstract *syntactic terms*, and not mereological parts of utterances (Werning, 2005); (b) syntax-driven composition (c) has immediate and predictable effects for updates of the model, as required by the Tarski–Montague–Partee view of the PSC.

Consider examples (1)–(4). We can ask what meanings would result, if each of these sentences were processed with a balance favoring, on average, the M-stream or the G-stream. The prediction is that *different meanings* will result in each case. Through the M-stream primarily, "John spilled the beans" means John revealed a secret, "The Beatles" refers to wax replicas of the musicians, "everyone" is every person in Bergen, and "the first case of AIDS" is the first person reported to have AIDS. If, however, the G-stream prevails on average, or at critical points during incremental interpretation, a different meaning would be assigned to at least some phrasal constituents: spilled beans from a container; the living members of The Beatles; everyone in the domain of discourse; the first person infected with HIV who had developed AIDS; and so on. Here, the intended meanings of (1)–(4) are those given by the M-stream. However, we can imagine contexts of use in which the appropriate meanings are, instead, those given by the G-stream. But what matters is that the "average" competent user of English is capable of generating (and understanding) both types of meanings, at least for some of (1)–(4). This validates dual-stream architectures, such as the one above. Moreover, trained linguists are generally capable of generating these different meanings at will. This suggests that switching from one stream to the other is or can be brought under cognitive control. Thus, one could be trained to "tune out" the G-stream and prioritize information provided by content words and the context to arrive at the appropriate interpretation, or alternatively to "tune out" the M-stream and prioritize grammatical information. The abilities to entertain different meanings for the same input strings, and to switch purposely between contextual M-stream and compositional G-stream processes, are accounted for by the parallel architecture: These abilities are intrinsic to the *competence of speakers* and to the *expertise of linguists*, respectively, and they in fact matter for a range of other linguistic practices (Borg, 2019).

For the present purposes, the most important feature of the parallel architecture is that it incorporates compositionality as a constraint on processing in the G-stream. It could be argued that, although this proposal *does* assign a definite functional role to compositionality, it does *not* specify what compositionality is nor what it implies. As noted in Section 1, the PSC does not do any work, if one does not also state what are the relevant "parts," meanings of the parts, and syntactic mode of composition. A psycholinguistic model needs a theory of grammar. Still, it is important to avoid the pitfalls of the methodological view of the PSC, where the question is *which theories*, not whether language and meaning, are compositional. So, one could design theories of lexical meaning and syntax that render the M-stream superfluous: Any computable meaning could be computed by the G-stream, if we regard the PSC as "carte blanche" for introducing complexity into the model's semantic or syntactic representations or operations (Baggio et al., 2012; Werning, 2005). Instead, the theory should serve an empirical strategy, where we try to determine what simple or complex expressions in sentences—beyond obvious candidates, such as various functors—are processed primarily by the G-stream, and under what conditions (see above). On that strategy, linguistic theory may provide initial impulse, suggest possible experimental designs, explanations of existing data, and so on, but it is largely through empirical—experimental and computational—research that one may draw a line, however squiggly, between what is compositional and what is not, that is, between meanings derived by relying *strictly* on lexical structures and syntax and meanings derived by harnessing instead the wider context and stored semantic information. Theories of lexical meaning and syntax in the brain, which will contribute to clarify the computational properties of G-stream processing and compositionality, will be written jointly by formal theorists and experimentalists. Here is one way this might happen.

Each processing phase in the parallel architecture may be related to modulations of different known components of the ERP (event-related brain potential; Fig. 1b). Interactions between bottom-up processes (thick arrows) and top-down processes (dotted arrows) can affect ERP amplitudes: For example, activation of the meaning of "socks" and relational processing in the context of the fragment "He spread the warm bread with ..." increases the amplitude of the N400 component (Baggio, 2018; Baggio & Hagoort, 2011; Kutas & Federmeier, 2011). Processing costs in the morphosyntactic (*m*), compositional (*c*), and interpretive (*i*) stages may increase the amplitudes of the LAN (left anterior negativity), P600, and SAN components (sustained anterior negativity; or Nref, referential negativity; Baggio, 2018), respectively. The timing of these effects is consistent with parallel processing in the two streams: the LAN and N400 occur approximately in the same window (300–500 ms after word onset), as do the P600 and the SAN (500–900 ms). The order of these effects, where the LAN precedes the P600, and the N400 precedes the SAN, is compatible with the two-phase, cascading nature of processing *within* each stream.

Research on the P600 can help answering questions on compositionality (Fritz & Baggio, 2020)—whether a particular expression, given the presence or absence of context, and for a specific experimental setting or task, is processed primarily by the G-stream. Through systematic study, we may hope to get to grips with the syntactic and semantic structures and operations most likely used by each stream. This is the cognitive neuroscience approach to the PSC, alternative and complementary to the linguistic approach, advocated in this paper (Section 1).

Cognitive neuroscience may provide evidence for or against the dual-stream model outlined above, also through functional anatomical data, beyond ERPs. One question is whether the M- and G-streams correspond to functionally and anatomically well-defined pathways—networks of regions where information flows following specific spatiotemporal patterns and contributes to specific linguistic computations—and whether, consequently, the composition phase in the G-stream may be associated with neural processing in a particular cortical region.

Attempts at localizing composition using magnetoencephalography (MEG) and fMRI have led to mixed results. MEG findings by Pylkkänen and colleagues initially seemed to suggest that the left anterior temporal lobe (LATL) instantiated a syntax-driven, logico-semantic form of composition (e.g., functional application; Bemis & Pylkkänen, 2011; Brennan & Pylkkänen, 2017; Westerlund, Kastner, Al Kaabi, & Pylkkänen, 2015; see also Matchin, Hammerly, & Lau, 2017 using fMRI). However, more recent studies have indicated that the LATL subserves non-syntactic, non-logico-semantic processes of *conceptual combination*. These LATL effects are sensitive to conceptual relations between the meanings of the "parts" of an NP, such as: the specificity of the N (e.g., relative to the less specific "spotted fish," which led to an LATL effect, the more specific "spotted trout" did not engage the LATL more than "trout" alone; Westerlund & Pylkkänen, 2014); the generality of the modifier (e.g., with "lamb stew" there was an LATL effect, but not with a more general modifier, like "meat" in "meat stew"; Zhang & Pylkkänen, 2015); the type of set-theoretic operation used (e.g., the LATL effect is stronger with intersective conjunctions, "The hearts were green and big," compared to collective conjunctions, "The hearts were small and big"; Poortman & Pylkkänen, 2016). LATL activity is unlikely to reflect composition (c), but rather early stages of relational semantic processing (r), during which local predication and modification processes are modulated conceptually (Baggio, 2018; Pylkkänen, 2016). Moreover, the LATL is not activated by suffixation (Flick et al., 2018), which shows it does not implement morphosyntactic G-stream processes (m).

From fMRI research we gather that the M-stream—relational (r) and interpretive (i) processing—corresponds to a broad network of regions, including the (anterior) left inferior frontal gyrus (aLIFG), the left posterior middle gyrus and superior temporal gyrus (pMSTG), the LATL, ventromedial PFC, and areas of the inferior parietal cortex (see Binder, Desai, Graves, & Conant, 2009 for context, Baggio, 2018 for a synthesis, and Molinaro, Paz-Alonso, Duñabeitia, & Carreiras, 2015 and Baggio et al., 2016 for two fMRI experiments). None of the regions in this network seem sensitive to syntactic structure or argument structure (Boylan, Trueswell, & Thompson-Schill, 2015; Fedorenko, Nieto-Castanon, & Kanwisher, 2012; Matchin, Liao, Gaston, & Lau, 2019; Schell, Zaccarella, & Friederici, 2017). A different set of regions, but still part of the classical perisylvian "language cortex," are involved in syntactic computations, and might thus constitute the functional anatomical basis of the G-stream. These comprise the middle and posterior portions of LIFG, the frontal operculum (FOP), the posterior superior temporal sulcus (pSTS), and the anterior insula (aINS; see, e.g., Pallier, Devauchelle, & Dehaene, 2011; Schell et al., 2017; Snijders et al., 2009; Zaccarella, Meyer, Makuuchi, & Friederici, 2017). It is still not clear to what extent this network carries out primarily or exclusively bottom-up computations (e.g., Merge; see Zaccarella et al., 2017) or also or primarily predictive, top-down processes (e.g., Matchin et al., 2017, Matchin, Brodbeck, Hammerly, & Lau, 2019; Matchin, Liao, Gaston, & Lau, 2019), as would be more consistent with the present model. In any case, this body of work provides suggestive evidence for distinct neuroanatomical pathways, subserving M- and G-stream processing. But much remains to be done to precisely demarcate the boundaries of composition(ality) in brain space and time (Olstad, Fritz, & Baggio, 2020).

5. Conclusions

The idea that the meaning of a complex expression is determined by the meanings of its constituents has been assumed in philosophy since Aristotle,[10] and it has been variously reformulated in the medieval and modern periods (Hodges, 2016), before taking its current shape in the works of Frege, Tarski, and Montague. It is only with Montague, however, that compositionality is applied to natural language. With him, compositionality migrates from logic to linguistics, where it has held the function of a yardstick against which to evaluate semantic analyses. Mindful of the long and tortuous history of compositionality, one may well be skeptical about its validity as an empirical hypothesis. Hence, critical approaches to the PSC seem justified.

In this paper, I have introduced one critical approach, inspired by ideas and results from the cognitive neuroscience of vision, memory, decision-making, and language. Let us summarize the main line of argument. (a) A cognitive neuroscience approach to compositionality complements the linguistic approach (Section 1). (b) Standard type-I (productivity, systematicity, etc.) and type-II (methodological and empirical) arguments are too weak: None of them can establish compositionality as a property of human language, following logically from given facts and assumptions. However, a new competence argument for compositionality may be given, where the PSC is an abstract, high-level, general statement on a specific aspect of semantic competence: the capacity to assign, to (some) complex expressions, meanings that are a function *only* of the meanings of the parts and of the expression's syntactic form (Section 2). (c) Neuroscience research points to representational and processing capacities of the brain consistent with new competence argument for compositionality (Section 3). (d) A model of language processing, with parallel streams for meaning and grammar, is described, which embodies these representational and processing capacities, and where compositionality is regarded as a constraint on syntax-driven composition (Section 4). This proposal reduces the scope of compositionality quite considerably: It is no longer a principle applying to language or to linguistic theory as a whole, but a computational constraint on one processing phase of four, in one processing stream of two, in the brain's language system.

Acknowledgments

I am grateful to two anonymous reviewers and to the participants in the "Compositionality in Brains and Machines" workshop at the Lorentz Center in Leiden and in the "International Brain and Syntax Think Tank" at Northwestern University for their helpful comments on the ideas presented here. This work was supported by the Research Council of Norway under project FRIHUMSAM 251219.

Notes

1. There is a question whether these are facts about language or linguistic (or semantic) competence. I will return to this issue in Section 2.
2. For a critical discussion of vacuity arguments, see Westerståhl (1998).
3. Such as satisfaction and truth, for example.
4. See Janssen (1997) for a discussion of three "general methods to obtain compositionality."
5. I use the collective term "entities" to refer to whatever it is that representations are *representations of*, in the language domain and beyond: for example, objects, events, rules, linguistic expressions, and so on.
6. Reverberi et al. (2012a) show that parietal cortex encodes rule and cue-identity information: Visual cues were used to initiate rule recall, so cue identity is relevant for assigning a "meaning" (i.e., a rule) to a cue. The parietal cortex may be involved in this "accessory task," mapping each cue to one rule, such as "If cue 1, then rule 3": "The representation of such rules, not being compositional at the conceptual level, should not be compositional at the neural level as well," they comment. This adds nuance to the conclusion that these two regions actually "represent the same information."
7. See Baggio (2018) for a discussion of this idea, also in relation to semantic compositionality (PSC). For applications of idea that parts should be "recoverable" from wholes, and that old parts should be "reusable" in new wholes, see Cole et al. (2011, 2013) and Duncan et al. (2017), among others.
8. In *decomposable* idioms, such as "to spill the beans," there is a one-to-one correspondence between the idiom's constituents and its meaning: "to spill" → to reveal, "the beans" → a secret. This is not the case in *non-decomposable* idioms, such as "to kick the bucket" (to die).

9. For example, see Heim & Kratzer (1998, p. 49): "We are adopting a view of the grammar as a whole on which syntax and semantics are independent modules. Each imposes its own constraints on the grammatical structures of the language."
10. See *On Interpretation*, VIII; from the Cooke and Tredennick (1955) translation: "If (. . .) one word has two meanings, which do not combine to make one, the affirmation itself is not one. If, for instance, you gave the name 'Garment' alike to a horse and a man, then it follows that 'Garment is white' would be not one but two affirmations, as would 'Garment is not white' be not one denial but two."

References

Aristotle, Tredennick, H., & Cooke, H. P. (1955). *The categories. On interpretation; Prior analytics*. Cambridge (MA), London (UK): Harvard University Press.

Baggio, G. (2018). *Meaning in the brain*. Cambridge, MA: MIT Press.

Baggio, G., Cherubini, P., Pischedda, D., Blumenthal, A., Haynes, J. D., & Reverberi, C. (2016). Multiple neural representations of elementary logical connectives. *NeuroImage*, *135*, 300–310.

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

Baggio, G., van Lambalgen, M., & Hagoort, P. (2012). The processing consequences of compositionality. In *The Oxford handbook of compositionality* (pp. 655–672). Oxford, UK: Oxford University Press.

Baroni, M. (2020). Linguistic generalization and compositionality in modern artificial neural networks. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20190307.

Battaglia, F. P., Borensztajn, G., & Bod, R. (2012). Structured cognition and neural systems: From rats to language. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1626–1639.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801–2814.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Borg, E. (2004). *Minimal semantics*. Oxford, UK: Oxford University Press.

Borg, E. (2012). *Pursuing meaning*. Oxford, UK: Oxford University Press.

Borg, E. (2019). Explanatory roles for minimal content. *Noûs*, *53*(3), 513–539.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73.

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, *78*, 130–141.

Braisby, N. (1998). Compositionality and the modelling of complex concepts. *Minds and Machines*, *8*(4), 479–507.

Brennan, J. R., & Pylkkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive Science*, *41*, 1515–1531.

Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson (Ed.), *Advances in psychology* (Vol. 77, pp. 217–240). Amsterdam: North-Holland.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

Carston, R. (2010). XIII—Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the Aristotelian Society*, *110*(3/3), 295–321.

Cheng, S., & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese*, *193*(5), 1345–1385.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.

Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., & Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*, 142.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348.

Cram, S., & Steedman, M. J. (1985). On not being led up the garden-path: The use of context by the psychological processor. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, UK: Cambridge University Press.

Dowty, D. (2007). Compositionality as an empirical problem. In C. Barker, P. Jacobson, & P. I. Jacobson (Eds.), *Direct compositionality* (Vol. 14, pp. 23–101). Oxford, UK: Oxford University Press.

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5295–5299.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, *17*(3), 449–472.

Eichenbaum, H., Schoenbaum, G., Young, B., & Bunsey, M. (1996). Functional organization of the hippocampal memory system. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13500–13507.

Fedorenko, E., Nieto-Castanon, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, *50*(4), 499–513.

Flick, G., Oseki, Y., Kaczmarek, A. R., Al Kaabi, M., Marantz, A., & Pylkkänen, L. (2018). Building words and phrases in the left temporal lobe. *Cortex*, *106*, 213–236.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*(2), 119.

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge, MA: MIT Press.

Fritz, I., & Baggio, G. (2020). Meaning composition in minimal phrasal contexts: Distinct ERP effects of intensionality and denotation. *Language, Cognition and Neuroscience*, *35*(10), 1295–1313.

Groenendijk, J. A. G., & Stokhof, M. J. B. (2004). Why compositionality? In J. Carlson & J. Pelletier (Eds.), *The partee effect* (pp. 83–106). Stanford, CA: CSLI Press.

Hannula, D. E., Ryan, J. D., & Warren, D. E. (2017). Beyond long-term declarative memory: Evaluating hippocampal contributions to unconscious memory expression, perception, and short-term retention. In D. E. Hannula & M. C. Duff (Eds.), *The hippocampus from cells to systems* (pp. 281–336). Cham, UK: Springer.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford, UK: Blackwell.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*(7), 523.

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 299–310). Amsterdam: Academic Press.

Hintikka, J., & Sandu, G. (1999). Tarski's guilty secret: Compositionality. In J. Wolenski & E. Köhler (Eds.), *Alfred Tarski and the Vienna Circle* (pp. 217–230). Dordrecht: Springer.

Hodges, W. (2016). Remarks on compositionality. In S. Abramsky, J. Kontinen, J. Väänänen, & H. Vollmer (Eds.), *Dependence logic* (pp. 99–107). Basel: Birkhäuser.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*, 2–22.

James, W. (1890). *The principles of psychology*. London: Macmillan.

Janssen, T. (1986). *Foundations and applications of Montague grammar, part 1*. Amsterdam: Center for Mathematics and Computer Science.

Janssen, T. M. (1997). Compositionality. In J. van Benthem & A. Ter Meulen (Eds.), *Handbook of logic and language* (pp. 417–473). Amsterdam: North-Holland. (with an appendix by B. Partee).

Keenan, E. L. (1978). On surface form and logical form. *Studies in the Linguistic Sciences—Urbana, IL*, *8* (2), 163–203.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Lewis, D. (1972). General semantics. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 169–218). Dordrecht: Springer.

Lewis, D. (1986). *On the plurality of worlds*. Oxford, UK: Blackwell.

Matchin, W., Brodbeck, C., Hammerly, C., & Lau, E. (2019). The temporal dynamics of structure and content in sentence comprehension: Evidence from fMRI-constrained MEG. *Human Brain Mapping*, *40* (2), 663–678.

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*, *88*, 106–123.

Matchin, W., Liao, C. H., Gaston, P., & Lau, E. (2019). Same words, different structures: An fMRI investigation of argument relations and the angular gyrus. *Neuropsychologia*, *125*, 116–128.

Michalon, O., & Baggio, G. (2019). Meaning-driven syntactic predictions in a parallel processing architecture: Theory and algorithmic modeling of ERP effects. *Neuropsychologia*, *131*, 171–183.

Molinaro, N., Paz-Alonso, P. M., Duñabeitia, J. A., & Carreiras, M. (2015). Combinatorial semantics strengthens angular-anterior temporal coupling. *Cortex*, *65*, 113–127.

Montague, R. (1970). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella società e nella tecnica* (pp. 189–223). Ivrea, Rome: Edizioni di Comunità.

Nefdt, R. M. (2020). A puzzle concerning compositionality in machines. *Minds and Machines*, *30*(1), 47–75.

Nordmann, E., Cleland, A. A., & Bull, R. (2013). Cat got your tongue? Using the tip-of-the-tongue state to investigate fixed expressions. *Cognitive Science*, *37*(8), 1553–1564.

Olstad, A. M. H., Fritz, I., & Baggio, G. (2020). Composition decomposed: Distinct neural mechanisms support processing of nouns in modification and predication contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2193–2206.

Pagin, P., & Westerståhl, D. (2010a). Compositionality I: Definitions and variants. *Philosophy Compass*, *5*(3), 250–264.

Pagin, P., & Westerståhl, D. (2010b). Compositionality II: Arguments and problems. *Philosophy Compass*, *5* (3), 265–282.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2522–2527.

Partee, B. H. (1984). Compositionality. In F. Landman & F. Veltman (Eds.), *Varieties of formal semantics* (pp. 281–312). Dordrecht: Foris.

Partee, B. H. (1995). Lexical semantics and compositionality. In D. Osherson (Ed.), *An invitation to cognitive science* (pp. 311–360). Cambridge, MA: MIT Press.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329–342.

Poortman, E. B., & Pylkkänen, L. (2016). Adjective conjunction as a window into the LATL's contribution to conceptual combination. *Brain and Language*, *160*, 50–60.

Pylkkänen, L. (2016). Composition of complex meaning: Interdisciplinary perspectives on the left anterior temporal lobe. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 621–631). Amsterdam: Academic Press.

Reverberi, C., Görgen, K., & Haynes, J. D. (2012a). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex*, *22*(6), 1237–1246.

Reverberi, C., Görgen, K., & Haynes, J. D. (2012b). Distributed representations of rule identity and rule order in human frontal cortex and striatum. *Journal of Neuroscience*, *32*(48), 17420–17430.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, *101*(3), 473–484.

Rubin, R. D., & Cohen, N. J. (2017). Memory, relational representations, and the long reach of the hippocampus. In D. E. Hannula & M. C. Duff (Eds.), *The Hippocampus from cells to systems* (pp. 337–366). Cham, UK: Springer.

Rubin, R., Schwarb, H., Lucas, H., Dulas, M., & Cohen, N. (2017). Dynamic hippocampal and prefrontal contributions to memory processes and representations blur the boundaries of traditional cognitive domains. *Brain Sciences*, *7*(7), 82.

Schell, M., Zaccarella, E., & Friederici, A. D. (2017). Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex*, *96*, 105–120.

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing verbs in German idioms: Evidence against the configuration hypothesis. *Metaphor and Symbol*, *22*(3), 213–231.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*(7), 1493–1503.

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161–184.

Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., & Nivre, J. (2008). The CoNLL 2008 shared task on joint parsing of syntactic and semantic dependencies. In A. Clark & K. Toutanova (Eds.), *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning* (pp. 159–177).

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*(5), 523–534.

Szabó, Z. G. (2000). Compositionality as supervenience. *Linguistics and Philosophy*, *23*, 475–505.

Tamber-Rosenau, B. J., & Marois, R. (2016). Central attention is serial, but midlevel and peripheral attention are parallel—A hypothesis. *Attention, Perception, & Psychophysics*, *78*(7), 1874–1888.

Tarski, A. (1956). *Logic, semantics, metamathematics*. Oxford, UK: Clarendon Press.

Titone, D. A., & Connine, C. M. (1999). On the compositional and non-compositional nature of idiomatic expressions. *Journal of Pragmatics*, *31*(12), 1655–1674.

Van Benthem, J. (1986). *Essays in logical semantics*. Dordrecht: Reidel.

Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, *32*(6), 939–984.

van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, in press (online first). https://doi.org/10.1177/1745691620970604

Vulchanova, M., Milburn, E., Vulchanov, V., & Baggio, G. (2019). Boon or Burden? The role of compositional meaning in figurative language processing and acquisition. *Journal of Logic, Language and Information*, *28*(2), 359–387.

Warden, M. R., & Miller, E. K. (2007). The representation of multiple objects in prefrontal neuronal delay activity. *Cerebral Cortex*, *17*(S1), i41–i50.

Warden, M. R., & Miller, E. K. (2010). Task-dependent changes in short-term memory in the prefrontal cortex. *Journal of Neuroscience*, *30*(47), 15801–15810.

Werning, M. (2005). Right and wrong reasons for compositionality. *The Compositionality of Meaning and Content: Foundational Issues*, *1*, 285–309.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language*, *141*, 124–134.

Westerlund M., & Pylkkänen L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59–70.

Westerståhl, D. (1998). On mathematical proofs of the vacuity of compositionality. *Linguistics and Philosophy*, *21*, 635–643.

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2017). Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex*, *27*(1), 411–421.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy*, *17*(4), 329–342.

Zhang, L., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, *111*, 228–240.