When Compositionality Fails to Predict Systematicity

Reinhard Blutner
Institute for Language, Logic and Computation (ILLC)
University of Amsterdam, Nieuwe Doelenstraat 15, 1012 CP Amsterdam, The Netherlands
blutner@uva.nl

Petra Hendriks
Center for Language and Cognition Groningen (CLCG)
University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands
P.Hendriks@let.rug.nl

Helen de Hoop
Department of Linguistics, CLS
University of Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands
H.deHoop@let.kun.nl

Oren Schwartz
Department of Cognitive Science
Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218
orens@cogsci.jhu.edu

Abstract

In this paper we argue that compositionality alone isn't sufficient to derive systematicity clauses for meanings (meaning = truth-functional content of an utterance). In a seminal paper by Fodor & Pylyshyn (1988) the 'systematicity' of adjectival modification has been derived from the stipulation that the syntactic operation of modification relates to the semantic operation of conjunction. We argue that this assumption fails already in the case of so-called 'absolute' adjectives. For instance, in order to determine the truth-functional content of a phrase like *red apple* it must be specified which parts of the apple are red, and this cannot be described by conjoining the adjective with the noun meaning.

Our answer to this and related puzzles is to take the meaning of the adjective as a function that applies to the meaning of the noun. Though compositional in the formal sense, this operation *fails* to predict any kind of systematicity for adjectival modification. In order to account for certain regularities we have to pose restrictions on the set of possible functions that count as adjective meanings. We are able to pose the right restrictions using a connectionist network that respects the role of encyclopedic knowledge. We argue that this model also is able to derive the prototype effects discussed in connection with adjectival modification.

Frege on Compositionality and Contextuality

For many semanticists, logicians and (generative) linguists the principle of compositionality of meaning has immediate appeal because it promises to have important empirical consequences that relate to the intuitive notion of systematicity: the ability to grasp clusters of new thoughts supposed some appropriate seed is given.[1] In this connection, Frege (1923) can be cited as it provides a clear illustration of the attitude under discussion:

> It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to the parts of a sentence, so that the structure of the sentence serves as the image of the structure of the thoughts (Frege (1923), translated by Geach and Stoothoff)

As Janssen (1997) points out, compositionality requires (a) that words in isolation have a meaning and (b) that from these meanings the meaning of a compound can be built. Interestingly, Frege did *not* believe in the idea of isolated word meaning when he wrote his *Grundlagen der Mathematik*: "One should ask for the meaning of a word only in the context of a sentence, and not in isolation"

[1] To be sure, our intuitive understanding of systematicity also includes significant effects of stochastic correlations.

(Frege, 1884, p. x). Following Janssen, this formulation of *contextuality* seems to disallow the meaning of words in isolation and is therefore incompatible with compositionality.

Though it is possible to reconcile contextuality and compositionality (Blutner, Hendriks, & de Hoop, 2003; Dummett, 1973) if a broad view of compositionality is taken, one important question remains: Assuming a proper way of filling in the compositionality schema, is it possible then to understand the regularities that lead to the clustering effect described above? The answer we will give is negative: compositionality alone is not sufficient. We need additional structural assumptions which can be abstracted from the nature of encyclopedic knowledge as represented in connectionist networks. This conclusion has important consequences for *cognitive architecture*. In order to account for the systematic nature of cognition it's not sufficient to stipulate a *formal* principle of compositionality. Moreover, general assumptions about connectionist networks are needed to overcome the shortcomings of classical architecture.

Compositionality versus Systematicity

The principle of compositionality of meaning can be seen as a formal *schema.* Usually, it is formulated as follows: The meaning of a compound expression is a function of the meanings of its parts (and of the syntactic rule by which they are combined). Its modern mathematical modeling is in terms of Universal Algebra where the compositionality relates to stating a homomorphism from the syntactic (term) algebra to the algebra of meanings. As Janssen (1997) points out, the principle of compositionality should not be considered an empirically verifiable restriction but a methodological principle that helps to find a proper design for syntax and semantics.

Close to the naïve conception of compositionality is that the syntactic algebra constitutes strings of words with the syntactic rule as rule of concatenation. Informally this constitutes a "way of *linking* or *ordering* successive constituents without altering them in any way as it forms the compound expression." (van Gelder, 1990). However, there are more powerful rules than just concatenation. Van Gelder (1990) collectively refers to such systems as constituting functional compositionality. He points out that the requisites of functional compositionality can be obtained whenever there are general, effective, and reliable processes for (a) producing an expression given its constituents, and (b) decomposing the expression back into those constituents.

In contrast to compositionality the notion of *systematicity* is much less clear. While some initial ideas can be traced back to Frege (see the above quote), the term and empirical hypothesis was introduced in Fodor & Pylyshyn (1988). Unfortunately, these authors didn't give a precise definition of the concept, or even an adequate description of the empirical phenomenon under discussion. In their discussion of systematicity, van Gelder & Niklasson (1994) offered the following scheme as a first approximation:

> *Systematicity according to Fodor & Pylyshyn* (*SFP)*: All cognitive systems (humans and other animals) are systematic, i.e., are such that their ability to do some things of a given cognitive type (including at least "thinking a thought" and making an inference) is intrinsically connected with their ability to do other, structurally related things of that type. (van Gelder & Niklasson, 1994)

Obviously, it has to be clarified what precisely counts as a structurally related thing (performance) for a given thing (performance) of a certain cognitive type in order to transmute this scheme into an empirical hypothesis. To be sure, if we restrict ourselves to particular domains, then it isn't difficult to fill out the scheme appropriately, and we can check whether the emerging hypothesis is a consequence of the principle of compositionality (and, perhaps, other classical assumptions) or not.

There is a large class of restricting adjectives that denote gradable properties, such as *tall, high, long, short, quick, intelligent*. It is well known that the applicability conditions of these adjectives vary depending upon the type of object to which they apply. What is high for a chair is not high for a tower and what is clever for a young child is not clever for an adult. On the other hand, there is the much more restricted class of so-called 'absolute' adjectives (such as *red*, *blue*, *spherical, quadratic, …*) for which it's claimed that their semantic values don't depend on the type of object to which they apply. Fodor & Pylyshyn (1988) consider the phenomenon of adjectival modification in connection with absolute adjectives and assume that the syntactical form '(adjective noun)$_N$' has the semantic force of a conjunction. In this case, the phenomenon of systematicity relates to clauses of the following kind:

(1) When an agent understands the expressions *brown triangle* and *black square*, she understands the expressions *brown square* and *black triangle* as well.[2]

According to Fodor & Pylyshyn, the systematicity of linguistic competence **derives** from the fact that the syntactic operation of modification relates to the semantic operation of conjunction (or intersection) in the discussed cases. Although they don't make any attempt to explicate this derivation, the underlying key ideas are clear enough.

First, a word of clarification is in order to avoid misunderstandings. The claim is *not* that systematicity gives us viable restrictions on the corresponding *extensions* of natural language expressions. For instance, if we know the *extensions* of BROWN$\cap$TRIANGLE and BLACK$\cap$SQUARE,

[2] We are not sure about the scope of systematicity clauses Fodor and Pylyshyn would generally permit. Szabó (2004) provides examples where our intuitions are more than a bit hazy: "But do all who understand *within an hour* and *without a watch* also understand *within a watch* and *without an hour*? And do all who understand *halfway closed* and *firmly believed* also understand *halfway believed* and *firmly closed*?". The question then is what kind of arguments permits or forbids such examples. In our view it has to do with the acquisition of encyclopedic knowledge.

then its normally not possible to derive the extensions of BROWN∩SQUARE and BLACK∩TRIANGLE[3].

According to the classical view, the phrase 'understanding natural language expressions' means more than just fixing a denotation. The following illustration may help to elucidate the point. We 'understand' a phrase as *brown triangle* if (i) we are able to grasp the corresponding conceptual representation (BROWN∩TRIANGLE), (ii) the corresponding conceptual components have an known truth functional impact[4] (i.e. they can be used to calculate the actual extensions if the relevant situational parameters are given), (iii) With the help of the involved logical operators it is possible to determine the truth conditions of the whole phrase under discussion.

Now we are ready to derive systematicity clauses such as (1) from the classical assumptions: When an agent understands the expressions *brown triangle* and *black square*, then she is able to construct (via compositionality) the conceptual representations BROWN∩TRIANGLE and BLACK∩SQUARE, respectively. Further, she knows the truth-conditional impact of the corresponding constituents. And she should be able to extract the lexicon entries *brown* → BROWN, *black* → BLACK, *triangle* → TRIANGLE, *square* → SQUARE. Using these entries and again the principle of compositionality, our agent is able to calculate the corresponding conceptual representations for *brown square* and *black triangle*. Moreover, she is able to calculate the truth-conditional impact of the corresponding conceptual representation using her knowledge about the truth-conditional impact of the conceptual constituents and the logical operation ∩. Hence, sharing the made assumptions, Fodor & Pylyshyn are right in assuming that classical architecture allows to derive the semantic systematicity of adjectival modification.

Quine and Lahav on Adjectival Modification

'Which parts of a apple have to be red in order to call it a red apple?' Obviously, an apple with yellow peel and red pulp doesn't count as a red apple under normal circumstances. Hence, the crucial insight is that the *truth-conditional content* determining what counts as a *red apple* is depending on selecting the proper parts of color assignment. For example, the truth conditions for *being a red apple* conforms to modifying the PEEL-PART of an apple (and not, saying, the PULP-PART).

In the previous section we have assumed that a big class of adjectives behaves intersectively. Now we see that this claim doesn't hold strictly. It appears valid only if we restrict the class of nouns considerably and consider nouns only that designate objects with a pure meronomy.[5] In general, however, all adjectives that are commonly considered as 'absolute' do not behave intersectively and show a dependence upon the objects class instead (like graded adjectives such as *big* and *tall*). Quine (1960) was the first who noted the contrast between *red apple* (red on the outside) and *pink grapefruit* (pink on the inside). In a similar vein, Lahav (1993) argues that an adjective such as *brown* doesn't make a simple and fixed contribution to any composite expression in which it appears:

> In order for a cow to be brown most of its body's surface should be brown, though not its udders, eyes, or internal organs. A brown crystal, on the other hand, needs to be brown both inside and outside. A brown book is brown if its cover, but not necessarily its inner pages, are mostly brown, while a newspaper is brown only if all its pages are brown. For a potato to be brown it needs to be brown only outside, ... (Lahav 1993: 76).

There are three consequences of this observation. First, we can no longer assume intersectivity, and Fodor & Pylyshyn's compositional analysis breaks down in the case under discussion. Second, we are not able to derive systematicity statements as (2) from the classical assumptions, since the intersection operation isn't applicable.

(2) When an agent understands the expressions *red apple* (RED PEEL) and *sweet grapefruit* (SWEET PULP), then it's likely that she understands the expressions *red grapefruit* (RED PULP) and *sweet apple* (SWEET PULP) as well.

Third, we conclude that encyclopedic knowledge is required to determine the truth conditional content of an utterance.[6] Possibly, the interaction of linguistic and encyclopedic knowledge may allow us to derive systematicity clauses as given in (2) – supposed the required knowledge about apples and grapefruits is available.

Prototypes and Adjectival Modification

Before we start to present our proposal for deriving systematicity we should mention two related problems that

---

[3] We use capitals for names of concepts and we use italics for the investigated natural language expressions (as in 'the noun *square* relates to the concept SQUARE which denotes the set of squares').

[4] This surely is the case for color concepts such as BROWN and form concepts such as TRIANGLE which correspond to perceptual classifiers. The same holds for a variety of other concepts that relate to a restricted class of predicates and relations (defined over a domain of objects, events, etc.)

[5] Two-dimensional mathematical objects such as squares and triangles seem to be of this type: they have a very poor meronomic structure. But even in these cases, adjectives do not always behave intersectively. Bosch (2002) presents an example of a triangle that has black edges and is white inside. It is possible to refer to this triangle as a *white triangle* in the context of an entirely black triangle and an entirely blue triangle, but as a black triangle in the context of a white triangle with red edges and a white triangle with blue edges.

[6] This point is generalized and carefully worked out in Relevance Theory (cf. Sperber & Wilson, 1986). In this theory the notion of explicature is used in order to fill the gap between formal, linguistic meaning and the propositional content (i.e., the explicit assumptions communicated by an utterance).

concern adjectival modification. The first problem goes back to Quine (1960) who noted the different colors denoted by *red* in *red apple* and *red hair*. A related observation is that the same color sometimes gets expressed in different ways, depending on what it contrasts with. For example, in Japanese, aka-zatoo 'brown sugar' (lit. 'red sugar') comes in the same range of colors as shira-miso, lit. 'white bean paste'. This replicates the earlier finding that adjectives as *red* don't behave as one-placed predicates that modify by conjoining with the noun's denotation. Hence, they don't make a simple and fixed contribution to any composite expression in which they appear. Instead, the actual color value deviates in a systematic way from the prototypical color value that can be assigned to the color adjective in isolation – in dependency on the conceptual properties of the modified noun.

The second problem concerns the typicality effects in adjective-noun combinations (e.g. Osherson & Smith, 1981; Smith, Osherson, Rips, & Keane, 1988). Most importantly, we have to account for the "conjunction effect" in this connection. To illustrate the effect consider the typicality of a particular red apple as an instance of the concepts APPLE and RED APPLE. The typical finding is that our red apple is judged more typical of RED APPLE than of APPLE. In case of "incompatible conjunctions" such as *striped apple* or *brown apple* the conjunction effect is greater than in "compatible conjunctions". Though the second problem doesn't relate to the truth-conditional impact of adjectival modification, we think a good model should account both for the truth-conditional peculiarities of adjectival modification and for the typicality effects.

Compositionality without Systematicity

There are at least two different theoretical possibilities to deal with the problems of adjectival modification. The first proposal conforms to the view of radical underspecification augmented with contextual enrichment (e.g. Blutner, 1998). In this view, every lexical unit determines an underspecified representation. The combinatorial system determines how lexical units are combined into larger units. Finally, there is a mechanism of contextual enrichment (pragmatic strengthening based on contextual and encyclopedic knowledge) which is controlled by factors of economy. As an example, we can take the meaning of the adjective *long* as λx LONG(x,X), denoting the class of objects that are long with regard to a comparison class which is indicated by the free variable X.

In a similar way we could introduce free variables addressing the different parts of an object that may be affected by the modifying adjective. However, this view leads to rather clumsy lexical entries. Worse, this theory does not really clarify how the border line between the underspecified representation and the contextual enrichment is ever to be determined[7], and it is difficult to see how the available mechanisms account for the prototype effects found in adjectival modification.

The second proposal takes Montague (1970) as its starting-point and considers the attributive use of adjectives as fundamental and accordingly does not treat adjectives as predicates, but rather as adnominal functors. Such functors turn the properties expressed by the noun into those expressed by the (adjective noun)$_N$ combination: (*adj noun*)$_N$ → ADJ'(NOUN'). To give an example, the functor associated with *red* may be defined disjunctively in the manner illustrated in (3):

(3) RED(X) means roughly the property
 (a) of having a red inner volume if X denotes fruits only the inside of which is edible
 (b) of having a red surface if X denotes fruits with edible outside
 (c) of having a functional part that is red if X denotes tools
 ...

The question now is: can we derive the intended systematicity clauses if we realize compositionality via adnominal functors. The answer is negative in the general case. Let's consider an example and assume our agent understands the phrase *red apple*. That means she is able to grasp the corresponding conceptual representation RED(APPLE), and, in order to determine the truth-functional impact she should also be able to calculate the value of the function RED for the argument APPLE. We assume the results is a COLOR attribute that is specifying the apple's PEEL-PART. Similarly, if our agent understands *sweet grapefruit* he grasps the conceptual representation SWEET(GRAPEFRUIT), and the value of the function SWEET for the argument GRAPEFRUIT is such that it specifies its PULP-PART.

The question now is, can we derive from these premises that our agent is able to determine the truth conditional impact for, let's say, *red grapefruit.* Compositionality may help her to determine the conceptual representation, RED(GRAPEFRUIT). But this is not enough information to determine the relevant truth-condition. She must also be able to calculate the value of the function RED for the argument GRAPEFRUIT. This value cannot be extrapolated from the value of the function RED for the argument APPLE. The intended solution is that this value corresponds to specifying the PULP-PART in case of our red grapefruit. However, the classical theory cannot explain it. Consequently, it cannot derive the systematicity clause given in (2).

The general conclusion is that compositionality is often very simple to realize if (higher order) functions are introduced. In many cases, this tool realizes generalizat-ions to the worst case (cf. the appendix in Jannsen 1997). Unfortunately, we cannot derive interesting systematicity clauses from this style of compositionality. The fixing of the functions for different values doesn't determine the complete functions without further information.

The previous discussion made clear that some kind of encyclopedic knowledge is required in order to fill in the systematicity gap. Using a Bayesian picture of the mental

[7] cf. Blutner, Hendriks, & de Hoop (2003)

encyclopedia it is plausible that at least certain *default* clauses of systematicity are derivable, which can be overwritten if more experience comes in.

Towards a Theory of Adjectival Modification

In this section we sketch a connectionist approach of overcoming the gap between compositionality and systematicity. It's our aim not only to model the truth-functional aspects of adjectival modification but likewise the typicality effects. On the first look, our model can be seen as a connectionist variant of the *selective modification model* of Smith et al. (1988). It shares with this model the (localist) attribute-value representation for the prototypes (apple, grapefruit, …) and for the relevant instances.

The architecture of a simplified variant of our model (not yet differentiating the parts of an object), is given in Figure 1. The activation values for the nodes at the conceptual layer conform to the numerical values assigned in the Smith et al. model.

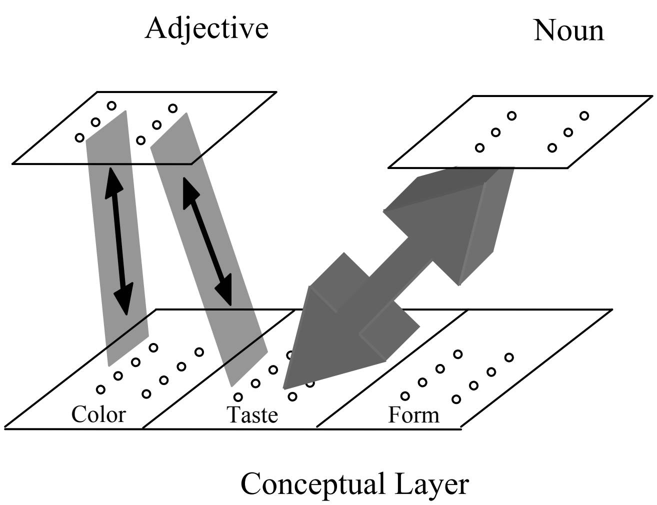

Figure 1: Architecture of a connectionist selective modification model (simplified). Inhibitory connections within the lower level aren't shown.

The upper layer of the model realizes localist representations of the adjectives and nouns, respectively (3 color adjectives, 3 taste adjectives, 6 nouns designating fruits). There are fixed bidirectional connections (maximum weight) between the nodes for the adjectives and the corresponding conceptual nodes at the lower layer (supposing a 1-1 correspondence between the adjectives and their localist attribute-value designations). The connection strength between the conceptual layer and the *noun* layer are learned using standard techniques of learning prototypes from exemplars (cf. McClelland & Rumelhart, 1988). For this purpose a representative sample of particular fruit exemplars was generated.

Though it's not possible here to give a detailed comparison between the connectionist modification model and the Smith et al. model, it should be mentioned that the basic findings are rather similar. Using Tversky's (1977) contrast rule (formulated for activation vectors)

(4) $\mathrm{sim}(\mathbf{s},\mathbf{t}) = \alpha\sum_i \min(s_i,t_i) - \beta\sum_i |s_i-t_i|$

we are able to demonstrate the conjunction effect. Take $\mathbf{s}_{apple}$ as realizing the activation vector for the conceptual layer supposing the noun *apple* is the input to the word layer, and take $\mathbf{s}_{red\ apple}$ as realizing the activation vector for the conceptual layer supposing the combination *red apple* inputs to the word layer. Further, let t_1 designate the activation vector for an instance of an red apple. Then our model correctly predicts the conjunction effect:

(5) $\mathrm{sim}(\mathbf{s}_{red\ apple,}\, t_1) > \mathrm{sim}(\mathbf{s}_{apple,}\, t_1)$

Moreover, we are able to account for the fact that the conjunction effect is greater in *incompatible conjunctions* than in *compatible conjunctions* (with t_2 an instance of a brown apple):

(6) $\mathrm{sim}(\mathbf{s}_{brown\ apple,}t_2) - \mathrm{sim}(\mathbf{s}_{apple,}t_2) >$
$\mathrm{sim}(\mathbf{s}_{red\ apple,}t_1) - \mathrm{sim}(\mathbf{s}_{\ apple,}t_1)$

In the next step of modeling we extend the model by distinguish different parts of the fruits (inside and outside). Since without further learning the model starts with a uniform weight matrix, the 'neutral' force of modification affects all parts uniformly. After learning, that the color of the outside of fruits is more discriminating than the color of its inside the learning mechanism correctly reproduces the expected sort of modification. Hence, the model behaves analoguous to Zwarts' (2003) "strongest meaning" model which starts with an initial default hypothesis which is subsequentially modified if more encyclopedic knowledge comes in. Concluding, we are able to derive systematicity clauses by making use of a connectionist modeling of the (raise of) encyclopedic knowledge.

Presently, we're working out a distributed version of the model making use of trichromatic representation of color and solving the arising binding problem by using tensor products. This extension will allow us to address more realistic data concerning the colors found in adjective noun combinations.

Finally, we want to stress that the present model is not intended to cover all aspects of compositionality and all aspects of systematicity. However, we think that the present discussion may help to understand (a) that compositionality per se is not a very restrictive empirical principle, (b) that systematicity of meaning has much to do with acquiring encyclopedic knowledge. Nevertheless, we believe that the principle of compositionality of meaning may be interesting – not as part of cognitive architecture, but as a consequence of evolutionary learning (e.g. Kirby, 2000). However, this is another story.

References

Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics, 15*, 115-162.

Blutner, R., Hendriks, P., & de Hoop, H. (2003). A new hypothesis on compositionality. In P. P. Slezak (Ed.), *Proceedings of the Joint International*

Conference on Cognitive Science. Sydney: ICCS/ASCS.
Bosch, P. (2002). *Explaining semantic productivity.* Paper presented at the Symposium on Logic and Creativity: Integrating Categorial Rules and Experience, Osnabrück.
Dummett, M. (Ed.). (1973). *Frege. Philosophy of Language*. London: Duckworth.
Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition, 28*, 3-71.
Frege, G. (1884). *Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl.* Breslau: W. Koebner.
Frege, G. (1923). Gedankengefüge. In G. Patzig (Ed.), *Frege, Logische Untersuchungen* (pp. 72-91). Göttingen 1976.
Janssen, T. M. V. (1997). Compositionality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of Logic and Language*. Amsterdam: Elsevier.
Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight & M. Studdert-Kennedy & J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 303-323). Cambridge: Cambridge University Press.
McClelland, J. L., & Rumelhart, D. E. (1988). A distributed model of human learning and memory. In J. L.McClelland & D. E. Rumelhart (Eds.), *Paralell Distributed Processing. Volume 2* (pp. 170-215). Cambridge: The MIT Press.
Montague, R. (1970). Universal grammar. *Theoria, 36*, 373-398.
Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9*, 35-58.
Smith, R., Osherson, D., Rips, L., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*.
Sperber, D., & Wilson, D. (1986). *Relevance*. Oxford: Basil Blackwell.
Szabó, Z. G. (2004). *Compositionality*. Stanford Encyclopedia of Philosophy. Available: http://plato.stanford.edu/entries/compositionality/
Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.
van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science, 14*, 355-384.
van Gelder, T., & Niklasson, L. (1994). Classicalism and cognitive architecture, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale NJ: Erlbaum.
Zwarts, J. (2003). Lexical Competition: 'Round' in English and Dutch. In P. Dekker & R. van Rooy (Eds.), *Proceedings of the Fourteenth Amsterdam Colloquium* (pp. 229-234). Amsterdam: ILLC.