

Optimality Theoretic Lexical Semantics

Lotte Hogeweg, Radboud University Nijmegen

1. Introduction

Optimality Theory (OT) has been applied to lexical semantics in several studies (e.g. Zwarts 2004, 2008, Zeevat 2002, Fong 2005, Hogeweg 2009a, b). The aim of the present article is twofold. Firstly, I want to explore the consequences of an OT approach to lexical semantics in more detail. Secondly, since the works mentioned only address functional items like prepositions and discourse markers, I will investigate the applicability of OT for the analysis of content words. This paper is organized as follows. In the next section, I will briefly introduce Optimality Theory. In section 3, I will address my first goal, which is to explore the consequences of an OT approach for the relation between words and meanings in more detail. I think there are three main consequences of such an approach and they will be discussed in section 3. In section 4, I will investigate whether OT is a suitable framework for analyzing the interpretation of content words.

2. Optimality Theory

Optimality Theory (OT) forms an important part of the Integrated Connectionist/Symbolic Cognitive Architecture (ISC) (Smolensky and Legendre 2006). ICS is a framework that integrates lower level connectionist representations and higher level symbolic representation. By doing so, symbolic theorizing has benefited from insights at the lower level of description. One of the most important insights was that networks can settle into a stable state through the interaction of conflicting forces (Soderstrom, Mathis and Smolensky 2006). Optimality Theory is based on this principle. In OT, linguistic knowledge is described as a system of ranked constraints. The constraints are ordered according to a strict priority ranking and they are potentially conflicting. A constraint may be violated to satisfy higher ranked constraints. OT hypothesizes that every language shares the same set of constraints. The difference between languages is due to a different ranking of those constraints.

OT specifies the relation between an input and an output. GEN (for generator) generates the possible output-candidates on the base of a given input. EVAL (for evaluator) evaluates the different candidates. The output that best satisfies the ranked constraints emerges as the optimal output for the given input. There are two types of constraints: faithfulness and markedness constraints. Faithfulness constraints order the output to be faithful to the input. Markedness constraints are solely concerned with the output. They indicate that an unmarked output is preferred over a marked output. To put it briefly, structures that are more complex are considered to be marked structures and structures that are

less complex or more natural are considered to be unmarked. Faithfulness to the input may sometimes require marked structures. Therefore, faithfulness and markedness constraints are potentially conflicting (Prince and Smolensky 1993/2004).

Phonology was the first area in linguistics to which Optimality Theory was applied. In phonology, constraints pertain to the relation between underlying form and surface form. Later the theory was also applied to syntax (Legendre, Grimshaw and Vikner 2001) and semantics (Hendriks and de Hoop 2001, and de Hoop and de Swart 2000). In OT semantics the input is an utterance and the output is an interpretation of that utterance.

OT has been applied to lexical semantics¹ in several studies. Fong (2003) shows that in Colloquial Singaporean English, the use of the adverb *already* is the result of the interaction of markedness and faithfulness constraints. Zwarts (2004) was one of the first to apply OT to the interpretation of lexical items, by giving an analysis of the interpretation of the preposition (*a*)*round*. Furthermore, in Hogeweg (2009a, b) the interpretation of the Dutch discourse particle *wel* is modeled in Optimality Theory. Common in these approaches is that a word is assumed to correspond to a fixed set of features. When a speaker wants to express a meaning, she compares the features in the input (the meaning she wants to express) to the bundles of features expressible by the lexicon of her language. Similarly, when a hearer interprets a word, she interprets the features that are stored for this word, provided that they are not in conflict with the (linguistic) context. As I mentioned in the introduction, I want to achieve two goals in this paper. Previous works on OT lexical semantics have addressed the adverb *already* (Fong 2003), several prepositions (Zwarts 2004, 2008) and various discourse markers (Zeevat 2002, Hogeweg 2009a, b). However, most work on lexical semantics focuses on content words like nouns, verbs and adjectives. In section 4, I will explore whether OT is a suitable framework for these types of lexical items as well. However, I will start in section 3 by exploring the consequences of an OT approach to lexical semantics. I believe that there are three important consequences of an OT view on lexical semantics: 1) words are not equal to concepts or contain concepts but a concept is the output of the process that takes a word(form) as an input or the input to the process has a word(form) as output 2) the meaning of words is overspecified in the lexicon 3) whether a concept is labeled by a particular word does not only depend on the stored information for that word but also on the stored information

¹ In this paper I use the term *lexical semantics* to refer to the relation between words and meanings. This includes both the selection as well as the interpretation of words.

of competing words, that is, competition is important. I will address each of the three points in the next section.

3. Optimality Theory and lexical semantics

3.1 Word-meaning relations as the outcome of optimization

An OT approach to lexical semantics entails that words do not have a one-to-one relation with meanings but the relation between words and meanings is the result of a process of optimization. A word is the output to the process that has a meaning as the input (production) or an input to the process that has a meaning as output (interpretation). In OT lexical semantics the input and output are both meanings. What is compared in the optimization of the interpretation of words is the fixed set of semantic features associated with the lexical item that forms the input and candidate interpretations for the word, which also consist of semantic features. In production, the input is a meaning the speaker wants to express and the candidate outputs are the sets of features that are conflated into words in the lexicon of the speaker. Hence, when I use the word *concept* I refer to a set of semantic features (or, as I will argue in section 4.2, attribute value structures). This contrasts with for example the use of the term concept by Osherson and Smith (1981) who argue that concepts that underlie kind terms such as *animal*, *tree*, *tool* are represented as a set of objects with information about the degree of prototypicality and the degree to which the object is characteristic for the concept. An object instances a concept to the extent that it is similar to the prototype of the concept. I assume that a word is not simply a label for a concept but that words are used for communication. A word can have a set of similar but not identical meanings. This is in line with the view put forward by Smolensky (1991) and criticized by Fodor and Pylyshyn (1988). Smolensky (1991) argues that the representation of the meaning of for example *a cup with coffee* varies according to the context in which it appears. According to Smolensky, we can depict the representation of *a cup with coffee* as the combination of certain semantic features like 'upright container', 'hot liquid', 'porcelain curved surface', 'burnt odor' etc. Critics of this view on representations would argue that it cannot be right because the representations of *cup without coffee* and *coffee* should be subtractable from the representation of *cup with coffee*. Now, Smolensky argues that we *can* subtract the representation of *coffee* from the representation of *cup with coffee*, only this will be a representation of *coffee* in a particular context. There is not one representation for coffee, but a collection of representations knit together by family resemblance. The particular representation that will emerge in a given situation is therefore context dependent. Nonetheless, *coffee* is a constituent of the representation of *cup with coffee*. However, this constituent relation is not part of the mechanism within the model.

Fodor and McLaughlin (1991) argue that systematicity requires *context-independent constituents*. In Smolensky's solution, the representation of *coffee* you get by subtracting it from the representation of *cup with coffee* is not a representation of *coffee* when it stands alone. The representation of coffee that you get from *cup with coffee* does not give the necessary conditions for being *coffee* because a representation of coffee in a can with coffee would yield a different set of features. And, Fodor and McLaughlin argue, it is not a sufficient set of features either. So, Fodor and McLaughlin wonder, what does make a representation a coffee-representation? There is no single vector that counts as *the* coffee-representation and therefore there is no vector that is a component of all the representations, which in a classical system would have *coffee* as a classical constituent. Fodor and McLaughlin suggest that Smolensky confuses being 'a representation of a cup with coffee with being a CUP WITH COFFEE representation':

"Espying some cup with coffee on a particular occasion, in a particular context, one might come to be in a mental state that represents it as having roughly the microfeatures that Smolensky lists. That mental state would then be a representation of a cup with coffee in this sense: there is a cup of coffee that it's a mental representation of. But it wouldn't of course, follow, that it's a CUP WITH COFFEE representation; and the mental representation of that cup of coffee might be quite different from the mental representation of the cup of coffee that you espied on some other occasion or in some other context. So, *which mental representation a cup of coffee gets is context dependent*, just as Smolensky says. But that doesn't give Smolensky what he needs to make representations themselves context dependent" (p. 342).

Smolensky argues that the semantic 'representation of a cup with coffee' can vary over contexts. Fodor and McLaughlin (1991) argue that the 'representation of a cup with coffee' that for example arises upon seeing one may be context dependent but this is not the type of representation that is part of the combinatorial system of language and thought. This latter type of representation, a CUP WITH COFFEE representation, is context independent.

In OT lexical semantics, the distinction between the two types of representations is the following: intentions (to express something) and interpretations are of the type 'representation of a cup with coffee'. This representation consists of a set of features which may vary across contexts. Words *are* linked to an invariable set of features. However, these features are not directly accessible to language users but they may surface in semantic representations that form the intentions or interpretations. As for the relation between words and meanings, Fodor and Lepore (2002) say: "we assume, for the

present discussion, that words express concepts, and that the content of a word is the content of the concept that it expresses". As said, I argue that words do not contain concepts but that concepts are the output to processes that take a word as their input or the input to processes that have a word as their output. The relevant question to ask then is not: what are the conditions for *being* COFFEE but what are the conditions for *calling* something coffee. The answer to this last question is that there are no necessary and sufficient conditions to label something as *coffee* but what matters is that the label *coffee* is better at expressing the intended meaning than the other available labels.

3.2 *Overspecified word meanings*

The studies mentioned in section 2 (except for Zwarts 2008, but see section 3.3) assume that a word is associated with a set of semantic features and that this meaning can be weakened by the context. In other words, the stored information about the meaning of words is overspecified. This contrasts with most current theories of lexical semantics in which it is argued that lexical representations are underspecified and may be strengthened by contextual information (e.g. Reyle 1993, Pustejovsky 1995, Blutner 1998, 2004). However, I think there are some good arguments for the position of overspecification, some of which are the consequence of an OT approach and some of which are independent. First the arguments that follow from OT. As outlined in section 2, there are two types of constraints in OT, faithfulness and markedness constraints. Zeevat (2000) argues that there is a fundamental asymmetry between OT syntax and semantics. Following Boersma (1998), he states that the production of speech and sentences stands under two opposing principles: expressiveness, which holds that the hearer of a sentence must be able to understand the message the speaker intended to express, and the principle of minimal effort or economy. The principles underlie the existence of respectively faithfulness and markedness constraints. Zeevat argues that the principle of minimizing effort is probably not at play in interpretation because if the hearer were to minimize her effort, she would run the risk of not finding the speakers intention. In other words, there does not seem to be a conflict between being faithful to the input to the maximal extent and the principle of economy. I therefore think that the only counterforce to faithfulness in interpretation is incompatibility, often instantiated in OT-semantics by a constraint FIT (interpretations should not conflict with the (linguistic) context, cf. Zwarts 2004, Hogeweg 2009a, b) or AVOID CONTRADICTION (cf. Hendriks & de Hoop 2001). If indeed the only contrary force to faithfulness is the avoidance of contradiction, as I assume, information about word meaning can only be deleted and not added, which means that we need to start with an overspecified representation. Another argument for an overspecified representation following from an

OT approach is given by de Hoop, Hendriks and Blutner (2007). They argue that with an OT-approach all possibly conflicting information sources are available from the start; hence we start with an overspecified meaning that can be weakened by contextual elements.

However, apart from the arguments following from an OT approach, I also think assuming overspecified lexical knowledge does a better job at explaining some phenomena and eventually may even be the most economical way of knowledge representation. It is often argued that words have an underspecified semantics and that the precise interpretation is accounted for by conceptual knowledge or commonsense knowledge (e.g. Bierwisch 1983). However, the nature of the mapping from semantic to conceptual knowledge is often not specified. An exception to this is Blutner (1998). Blutner argues for radical underspecification of lexical knowledge and that the lacking information is filled in by means of abduction rules. For example, upon hearing *red apple*, abduction rules determine the 'price' of interpreting *red* as referring to the color of the peel of the apple as opposed to the color of the stem or the pulp. The price is determined based on a Horn Clause Knowledge base containing clauses of the form $p_1, \dots, p_n \rightarrow q$, where the literals p_i in the antecedent are annotated with weights. This approach entails that for every word, it is specified what parts the concept it refers to consists of and how likely it is that it is this part that is being modified by an adjective. This is a lot of knowledge to represent but if you want to give an explanation of the complete interpretation process, you will eventually need this information. However, instead of saying that this information lies outside the scope of semantics I would like to argue that it is part of our lexical knowledge. Knowledge about concepts (of which I suspect everybody agrees that people possess it) contributes to compositional interpretation and can explain phenomena for which otherwise extra semantic knowledge has to be assumed, which would mean that the same information is represented twice. This aspect of my proposal is in line with Jackendoff's Conceptual Semantics (e.g. 2006, in press), which assumes that words are directly linked to conceptual knowledge.

With respect to overspecification, a note should be made on the role of strength and relevance in interpretation. The original idea of overspecification in OT lexical semantics as proposed by Zwarts (2004) was based on the Strongest Meaning Hypothesis (SMH) by Dalrymple, Kanazawa, Kim, Mchombo & Peters (1994). The SMH was proposed to handle the interpretation of reciprocals. However, the notion strength has also been argued to play a role in the interpretation of linguistic entities beyond the level of words, for example in the interpretation of plural predicates (Winter 2001). Zeevat analyzes the preference for global accommodation with presuppositions as a result of strength (2002) or maximization of relevance (2006). Van Rooij (2003) disentangles strength and relevance and argues that

the latter instead of the first explains the non-exhaustive reading of an answer like *around the corner* to questions such as *Where can I buy an Italian news paper?*. These previous analyses raise the question whether the proposal as outlined in this paper is compatible with the accounts of presupposition and non-exhaustive answers in terms of strength and relevance. An analogy like the following suggests itself. Strength causes the activation of all information associated with the form presented. Information can be canceled due to failing relevance in a way similar to how FIT cancels conflicting features from a representation. However, one of the improvements of the current proposal with respect to previous works is the precise determination of the circumstances in which FIT is violated (see section 4.2). A more precise investigation of the abovementioned phenomena in relation with FIT and faithfulness constraints should give insight into possibility to analyze them in such terms.

3.3 Competition

Competition is a crucial factor in OT, and hence in OT lexical semantics. This also follows from the view on the relation between words and meanings outlined above. The interpretation of a word depends on the context in which it is used. One and the same word may have a different contribution to the meaning of a sentence dependent on the context. Candidate interpretations therefore have to compete for becoming the optimal output during the process of interpretation. Similarly, during production candidate forms, or rather the bundles of features the forms express, compete for being chosen to express the intended meaning. To illustrate this latter process and to show the importance of competition I would like to briefly discuss the analysis of Zwarts (2008) of the selection of prepositions. Zwarts models the mapping from spatial meaning to prepositions in an Optimality Theoretic framework. In his model, the input is a bundle of features and the candidates are formed by a set of words from a particular language that express one or more of the input features. Every semantic feature corresponds to a faithfulness constraint $FAITH(F_i)$. This constraint is violated if the relevant feature is part of the input but is not reflected in the output. Say, the input is the set of features F and G: $\{F, G\}$. There may be three relevant candidate words for this input, namely a word that expresses F ($WORD_F$), a word that expresses G ($WORD_G$) and a word that expresses F and G ($WORD_{F,G}$). However, Zwarts assumes that prepositions maximally express one feature. The choice is therefore between ($WORD_F$) and ($WORD_G$). There are two constraints relevant to this situation, $FAITH(F)$ and $FAITH(G)$. The question is which faithfulness constraint is ranked higher, that is, which feature is more important to express. Tableau 1 illustrates the optimization process in a situation in which $FAITH(F)$ is ranked higher than $FAITH(G)$.

F, G	FAITH(F)	FAITH(G)
☞ (WORD _F)		*
(WORD _G)	*	

Tableau 1: lexical optimization

In tableau 1, the input consists of the features F and G and there are two relevant candidates (WORD_F) and (WORD_G). Since FAITH(F) is ranked above FAITH(G), (WORD_F) is the optimal expression for the input. Zwarts (2008) applies this model to the production of prepositions.

Prepositions typically express a relation between a Figure and a Ground. Zwarts argues that the English preposition *in* is associated with the semantic feature CONTAINMENT (CONT), which means that the Figure is contained by the Ground. The preposition *on* is associated with the feature SUPPORT (SUP), which means that the Figure is supported by the Ground. There are relations between a Figure and a Ground that can be characterized as pure CONTAINMENT, such as a fish swimming in the water, or as pure SUPPORT, such as a cup being on the table. However, Zwarts focuses on the situations in which both CONTAINMENT and SUPPORT apply. In a situation of a pen being in a box, for example, the pen is both contained and supported by the box. The example already shows that such a relation between Figure and Ground is expressed by the preposition *in*. Zwarts argues that this is due to the fact that the relation CONTAINMENT takes priority over the relation SUPPORT. He therefore concludes that FAITH(CONT) is ranked higher than FAITH(SUPP). This yields the following tableau.

[CONT, SUPP]	FAITH(CONT)	FAITH(SUPP)
☞ in _{cont}		*
on _{supp}	*	

Tableau 2: *in* versus *on*

The input to tableau 2 consists of the features CONTAINMENT and SUPPORT. The two relevant candidates are *in*, which expresses the feature CONTAINMENT, and *on* which expresses the feature SUPPORT. Both candidates leave one feature unexpressed but since it is more important to be faithful to the feature CONTAINMENT than it is to be faithful to SUPPORT, *in* is optimal.

A similar analysis is provided for the preposition *on* versus *over/above*. When a Figure is supported by a Ground, the Figure is typically also higher than the ground. This relation is expressed by the feature SUPERIOR (SUP). This results in the feature set {SUPPORT, SUPERIOR}. When only the feature

SUPPORT applies (as in the example of a picture on the wall), or when both features apply (when for example a vase is on a table) *on* is used while in situations where only SUPERIOR applies (as in the example of a bird flying over a house), *above* or *over* is used. This indicates that FAITH(SUPP) is ranked higher than FAITH(SUPER).

In addition, a similar analysis is provided for the prepositions *around* and *over*. The preposition *over*, in for example *the bird flew over the yard*, describes that an object moves along a path that is located above the Ground. This use of *over* expresses the features {PATH, SUPERIOR}. When both the features SUPERIOR and CONVEX are part of the input, in addition to the feature PATH, for example in a situation where a someone climbs over a wall, the proposition *over* will be chosen which shows that FAITH(SUPERIOR) is ranked higher than FAITH(CONVEX).

The priorities distinguished by Zwarts (2008) lead to the following ranking of constraints:

- (1) FAITH(CONTAINMENT) >> FAITH(SUPPORT) >> FAITH(SUPERIOR) >>
FAITH(CONVEX)

Zwarts (2008) offers the following explanation for the hierarchy. Usually, when an object is held by a container, this object is also supported by it. Therefore, situations that involve CONTAINMENT typically also involve SUPPORT. Similarly, situations that involve SUPPORT typically involve the feature SUPERIOR, since an object which is supported by another object, say a vase by a table, is usually also superior to it. Furthermore, paths with the feature SUPERIOR typically also have the feature CONVEX. When a cat jumps over a hedge, for example, it follows a path that is curved around the hedge. So, *in* expresses CONTAINMENT and although CONTAINMENT typically brings along the feature SUPPORT, Zwarts argues that this feature cannot be an inherent lexical feature of the form *in*, because it is not applicable to every situation in which *in* is used. However, as discussed in section 3.2, due to a constraint FIT or AVOID CONTRADICTION, features that are associated with a form can be dismissed by the context. So, the fact that SUPPORT is not part of every interpretation of *in*, does not necessarily mean that it is not a semantic feature of this preposition. In line with the other OT approaches to lexical semantics, I propose that also in this analysis a form is associated with a meaning (a set of features) which can be weakened by the context. This way the entailment relation between the features is reflected in the candidates rather than in the constraints. The advantage of this approach is that the choice for one preposition over the other can be explained by a more general principle of faithfulness to the input. Say, *in* is associated with both the features CONTAINMENT and SUPPORT and *on* is associated with SUPPORT. When the input is the set of features {CONT, SUPP} the preposition *in* is optimal since it satisfies both FAITH(CONT) and FAITH(SUPP). The

hierarchy in (1) can then be explained by the existence of an entailment relation between the features. If the property CONTAINMENT is absent in the output, this entails a violation of more features than when the property SUPPORT is violated, since this is a bigger violation of the faithfulness constraint MAX, which requires all input segments to have output correspondence. The reformulation of the relation between the features and constraints still yields the right result. In a situation in which the input is the set of features {CONT, SUPP} the preposition *in* is optimal since it satisfies both FAITH(CONT) and FAITH(SUPP).

[CONT, SUPP]	FAITH(CONT)	FAITH(SUPP)
☞ <i>in</i> _{cont, supp}		
☞ <i>on</i> _{supp}	*	

Tableau 3: *in* versus *on* 2

In tableau 3, the input is the feature set {CONT, SUPP}. The relevant candidates are *on*, which expresses SUPPORT and *in*, which now expresses both CONTAINMENT and SUPPORT and thereby satisfies both constraints. This tableau yields the right result, similar to Zwarts' analysis. The difference is that the result now reflects the explanation given by Zwarts concerning the typical entailment relations between features. Furthermore, the result is explained by a more general principle of faithfulness to the input, since the choice for *in* results in faithfulness to two features while the choice for *on* results in faithfulness to only one feature. If the input is the feature SUPPORT, both candidates are equally well since they both satisfy FAITH(SUPP) and FAITH(CONT) does not apply, as is illustrated by tableau 4.

[SUPP]	FAITH(CONT)	FAITH(SUPP)
☞ <i>in</i> _{cont, supp}		
☞ <i>on</i> _{supp}		

Tableau 4: *in* versus *on* 3

This incorrect outcome can be helped if we use the constraint DEP, which is well-known in phonology.

DEP: Do not express features not present in the input.

The constraint DEP is violated by candidates that express a feature that is not part of the input. If we apply this constraint to the situation in tableau 4, this would yield the correct result, namely that *on* is used when only SUPPORT is in the input, as is represented in tableau 5.

[SUPP]	FAITH(CONT)	FAITH(SUPP)	DEP
$in_{cont, supp}$			*
on_{supp}			

Tableau 5: *in* versus *on* 4

If only the feature CONTAINMENT is in the input, *in* is still optimal, which is illustrated in tableau 6.

[CONT]	FAITH(CONT)	FAITH(SUPP)	DEP
$in_{cont, supp}$			*
on_{supp}	*		*

Tableau 6: *in* versus *on* 5

In tableau 6, the optimal candidate violates DEP, but because there is no candidate available that has only the inherent feature CONTAINMENT, *in* is still optimal. In a situation where only the feature SUPPORT applies, for example in a situation where a vase is on a table, *in* is not optimal because it violates DEP. In a situation where only CONTAINMENT applies, for example when a fish is swimming in the water, *in* also violates DEP but is nonetheless optimal because in this case there is no better candidate available. There is no form available that only expresses the feature CONTAINMENT. Therefore the optimal choice is the item that expresses CONTAINMENT and SUPPORT. The analysis by Zwarts (2004), in this slightly altered version, shows that competition between forms is crucial in the choice for a lexical item. Words do not have to be perfect, as long as they are optimal, that is, as long as there is no better form available to express the intended meaning.

As I mentioned, a number of studies have taken an OT approach to lexical semantics. The studies all addressed function words. The majority of lexical semantic theories however, primarily addresses content words. In the next section I will make a start with answering the question whether OT is suitable for the analysis of content words.

4. OT and the semantics of content words

As I explained earlier, previous OT work in lexical semantics targeted highly grammaticalized items like prepositions, aspectual markers and discourse markers. Needless to say, a theory of lexical semantics should be able to account for all types of words. Literature on lexical semantics usually focuses on content words like verbs, nouns and adjectives. In this section I will explore whether OT is a suitable framework for handling this group of words as well. I will focus on the semantics of nouns and noun

adjective combinations. Of course, a complete theory of lexical semantics should be able to account for the full range of lexical items including verbs. This goes beyond the scope of the paper however. The consequence of this is that the claims I can make about my proposal are limited, but I hope that I can show that it is an enterprise worth pursuing.

4.1 Content words and overspecification

In the studies discussed so far, the meaning of words consists of semantic features, as for example the features CONTAINMENT and SUPPORT in Zwarts (2008). For content words this approach may seem more problematic at first sight than for the highly grammaticalized items addressed so far. An apple, just to name something, can be green, red, or another color. Which of those features is included in the representation of apple? A lexical semantics based on semantic features was earlier proposed by Katz & Fodor (1963). They argue that meanings are bundles of features and that the meaning of a combination of an adjective and a noun is the sum of the meaning of its parts. Kamp and Partee (1995) show that this approach only works for the so-called intersective adjectives. An adjective like *carnivorous* is intersective in that for any N it holds that 'carnivorous N' = 'carnivorous \cap N'. An adjective like *skilful* is not intersective for a skilful surgeon is not necessarily a skilful pianist. *Skilful* is a subsective adjective because for every N it holds that 'skilful N' \subseteq 'N'. Adjectives like *former* and *alleged* are neither intersective nor subsective because a former senator is not a member of the set of senators. Nonsubsective adjectives may be plain nonsubsective (no entailments at all, such as *alleged*, *arguably* etc.) or privative (entailing a negation of the noun property, such as *past*, *would-be* etc.). I will argue that the problem addressed by Kamp and Partee (1995) with the semantic feature-approach by Katz and Fodor (1963) fails to exist when instead of determining a defining set of features, we assume overrepresentation of features. In the next section I will first argue that instead of a flat feature representation, we need attribute value structures to represent the meaning of (content) words. I will argue for this by means of an analysis of the interpretation of color-denoting adjectives.

4.2 Attribute value structure

A case in point in the discussion about compositionality has always been the context dependence of color-denoting adjectives. While the adjectives seem to behave intersectively at first sight, they actually do not (e.g. Quine 1960, Lahav 1989, Blutner 2004). For example, red in *a* denotes the color of the peel of the apple while red in *b* denotes the color of the inside of the grapefruit.

a. Red apple

b. Red grapefruit

While some authors argue that the systematic behavior of color-denoting adjectives can be used as an argument for the principle of compositionality (e.g. Fodor and Pylyshyn 1988), others take the flexibility of the meaning of the adjectives as an argument against compositionality (Lahav 1989). Montague (1970), Keenan (1974) and Kamp (1975) propose a solution in which adjectives are adnominal functors which (for example) turn the properties expressed by apple into those expressed by red apple. This view entails that such functors have to be defined disjunctively (Blutner 2004):

RED(X) means roughly the property

- o Of having a red inner volume if X denotes fruits only the inside of which is edible
- o Of having a red surface if X denotes fruits with edible outside
- o Of having a functional part that is red if X denotes tools

An obvious problem of such an approach is that one cannot produce or interpret *red* as a predicate of nouns not in the list. Another approach to the problem is to assume a two level semantics whereby the lexical information is underspecified, for example carrying variables. (Bierwisch and Schreuder 1992, Blutner 2004). However, most authors agree that the interpretation of an utterance includes what is mostly considered “conceptual” knowledge:

“Bierwisch and Schreuder (1992): “... the concept structure CS, in terms of which the **actual** interpretation of linguistic expressions is specified...”(p. 32, boldface mine).

“Blutner (2004): “Of course it is not sufficient to postulate underspecified lexical representations and to indicate what the sets of semantically possible specifications of the variables are. In order to grasp natural language interpretation it is also required to provide a restrictive account explaining how the free variables are instantiated in the appropriate way. Obviously, such a mechanism has to take into consideration various aspects of world and discourse knowledge” (p. 5).

As I argued before, if we leave aspects of the interpretation to conceptual knowledge or world knowledge without specifying how this is done, we are pushing the problem down a level. As I already suggested in section 3, an overspecified lexical representation includes what is usually considered

conceptual, encyclopedic or commonsense knowledge. This means that lexical knowledge specifies, for example, that an apple has a stem, pulp, a peel etc. It also specifies that apples, or to be more precise, the peel of an apple can be red or green. So, an apple could be represented as {pulp, stem, peel, red V green}. However, this does not specify that it is the peel that is green or red and there is for example no way to indicate that the stem is most probably brown. In other words, there needs to be more structure in the set of features, in the form of recursive attribute value pairs. Barsalou (1992) argues that frames as recursive attributevalue s tructures provide the fundamental representation of human cognition. Rather than consisting of a feature set at a flat level of analysis, concepts are formed by attribute-value sets, with some characteristics (values) being instances of other characteristics (attributes). For example: blue and green are values of the attribute color, swim and fly are values of locomotion. Whereas the theory of frames as proposed by Barsalou was not formalized, nowadays formal theories of frames are available. Petersen (2007) for example, provides a formal theory of frames as connected directed rooted graphs with labeled nodes and arcs. The attributes in the frames assign unique values to concepts and are therefore functional relations. The values in frames can be atomic or composite (in which case the value is further specified by attributes of its own). Furthermore, they can be more or less specific. For example, the value of the attribute 'color' can be 'color' or 'red'. Frames come with appropriateness conditions restricting the possible values for an attribute. Appropriateness is determined based on a type signature and the condition that the possible values of an attribute are subtypes (e.g. red, blue, indigo) of the type corresponding to the attribute (color). Concepts can also have more specific appropriate conditions. Since an apple is always round, for example, the appropriateness condition shape:shape for apple is tightened up to shape:round. Figure 1 gives an example of a frame of a lolly pop, based on Petersen (2007).

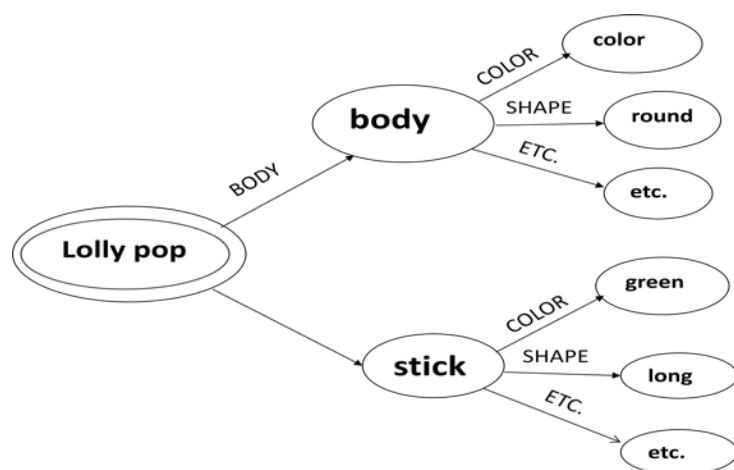


Figure 1: lolly pop frame

For the purpose of this article I will disregard most of the details of the formalization of frames and I will represent them as follows in (2):

- (2) $\lambda x[\text{lolly pop}(x) \wedge \text{body of}(\text{body}, x) \wedge \text{color of}(\text{green}, \text{body}) \wedge \text{shape of}(\text{round}, \text{body}) \wedge \text{taste of}(\text{apple}, \text{body}) \wedge \text{etc.}]$

Frames are compatible with connectionist networks, in which OT has its base (cf. Rumelhart, Smolensky, McClelland & Hinton 1986, Blutner, Hendriks, de Hoop, and Schwarz 2004). It should be noted however, that the representation of concepts as frames function at a more abstract level than representations in terms of units and connections. A connectionist network representing the concept 'lollypop' would not have a particular unit representing the whole concept, connected to a unit representing the body, connected to a unit representing the color of the body, etc. The node representing the whole concept 'lollypop' in this case represents a set of units including the same units that form the body node. For more information about the relation between connectionist networks and frames or schemata see Rumelhart, Smolensky, McClelland & Hinton (1986).

With respect to adjective noun combinations, Petersen, Fleischhauer, Beseoglu and Bücken (2008) argue that in frame theory adjectives could in general be represented as in Figure 2.



Figure 2: adjective

The semantic contribution of the adjective to the compound is twofold: (1) it selects an attribute and (2) it modifies the attribute value (Petersen et al. 2008). As such, the merger of the two frames is similar to the mechanism of selective binding as proposed by Pustejovsky (1995). Pustejovsky (1995) introduces an extensive theory of lexical semantics called the Generative Lexicon (GL). GL entails a core set of word senses, with a greater internal structure than usually assumed, which can generate a larger set of word senses when combined with other words in phrases and clauses. The semantics of a lexical item α can be defined as a structure consisting of four elements:

$$\alpha = \langle A, E, Q, I \rangle$$

where A is the argument structure, E is the specification of the event type, Q provides the binding of these two parameters in the qualia structure and I is an embedding transformation, placing α in a type lattice, determining what information is inheritable from the global lexical structure. For the present purpose I will focus on the Qualia structure. The Qualia Structure specifies four basic aspects of a word's meaning, based on the 4 modes of explanation by Aristotle:

- Constitutive: the relation between an object and its constituent parts. Possible values that each role may assume are material, weight, parts and component elements.
- Formal: that which distinguishes the object within a larger domain. Possible values: orientation, magnitude, shape, dimensionality, color, position.
- Telic: its purpose and function. Possible values: purpose that an agent has in performing an act, built-in function or aim which specifies certain activities.
- Agentive: factors involving its origin or "bringing it about". Possible values: creator, artifact, natural kind, causal chain.

Every category expresses a qualia structure but not all lexical items carry a value for each qualia role. Now, consider example (3a) and (3b).

- (3) a We will need a fast boat
 b John is a fast typist

Pustejovsky argues that if we treat the adjective *fast* as an intersective adjective we get: $\lambda x[\text{typist}'(x) \wedge \text{fast}'(x)]$. But this does not give us the right interpretation: 'John is a typist who is fast at typing'. In order to get this, the adjective is able to make available a selective interpretation of an event expression contained in the qualia of the head noun. What makes this possible is the generative mechanism called selective binding:

Selective binding

If α is of type $\langle a, a \rangle$, β is of type b , and the qualia structure of β , $QS\beta$, has quale, q of type a , then $\alpha\beta$ is of type b , where $\{\alpha\beta\} = \beta \cap \alpha(q\beta)$.

The denotation of $\{\alpha\beta\}$ is the noun intersected with the adjective applied to the relevant quale of the noun. The adjective can be seen as a function applied to a particular quale within the N that it is composed with. I propose that selective binding is the preferred way of combining an adjective and a noun. This is also argued for by Smith, Osherson, Rips and Keane in their selective modification model and is in line with the Head primacy principle, as proposed by Kamp and Partee (1995, p. 161):

“In a modifier-head structure, the head is interpreted relative to the context of the whole constituent, and the modifier is interpreted relative to the local context created by the former context created by the head.”

This principle is extended by Winter (2010) to the plug-in principle:

Plug-in principle

When a typicality-based representation T of an expression $E1$ takes part in the meaning composition of a complex expression $E1-E2$, contextual parameters of $E2$ are most sensitive to (“plugged-into”) the parallel features in T , if such features exist.

Winter argues that with assuming this principle, contextual parameters actively participate in the composition of meaning, rather than only playing a role in the use of such meanings. The preference for selective binding can be implemented as a constraint against non-selective binding *Non-selective Binding (*NB). Say, (part of the) the representation of apple is: $\lambda x[\text{apple}(x) \wedge \text{peel of } (\text{peel}, x) \wedge \text{color of } (\text{green} \vee \text{red}, \text{peel}) \wedge \text{stem of } (\text{stem}, x) \wedge \text{color of } (\text{brown}, \text{stem}) \wedge \text{etc.}]$ and the representation of red is $\lambda y[\text{red}(y)]$ In addition to the constraint *NB, we have the constraints FAITH which tells candidates to be faithful to all features in the input and the constraint FIT which penalizes candidates with (internal) inconsistencies, which is the case in three situations: 1: when there are two or more values for the same attribute, and 2: when a value or attribute is of the wrong type as determined by a type hierarchy, for example the value warm for the attribute color or the attribute color for the type smell, and 3: when a value of an attribute is incongruent with an inherited value. This leads to the following tableau:

Input: $\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red } \vee \text{ green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$ $\lambda y[\text{red}(y)]$	FIT	*NB	FAITH
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red } \vee \text{ green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown } \wedge \text{ red, stem)} \wedge \text{etc.}]$	*		
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red } \vee \text{ green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{color of (red, x)} \wedge \text{etc.}]$	**	*	
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red } \vee \text{ green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (red, stem)} \wedge \text{etc.}]$			*
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$			
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (red, stem)} \wedge \text{color of (red, x)} \wedge \text{etc.}]$		*	**

Tableau 6: interpretation of *apple*

The first candidate in which the adjective *red* is interpreted as referring to the color of the stem of the apple violates FIT because it contains the conflicting information that the stem of the apple is both brown and red. The second candidate violates the constraint *NB because the adjective modifies the complete concept ‘apple’ rather than an appropriate part with a color slot. The same holds for the last candidate with the difference that the stem is red as well. The third candidate violates FAITH because it does not reflect the information present in the input that the color of the stem is brown. The fourth candidate does not violate any constraint because it is faithful to the information in the input, the adjective is selectively bound and it contains no inconsistencies. However, the current set of constraints wrongfully predicts that the interpretation of *brown apple* would be an apple with a brown stem. Furthermore, there is no way of predicting the interpretation of for example *lilac apple*. In order to get this, we need two things. First, we need a constraint that says that the information that is provided by the adjective overrides the information that is part of the noun. We interpret the phrase *lilac apple* as referring to an apple that is lilac and not red or green. Such a constraint was proposed by Kamp and Partee (1995) as the Non-vacuity Principle (NVP): ‘In any given context, try to interpret any predicate so that both its positive and negative extension are non-empty’ (p. 161). If we interpret *lilac apple* as a red or green apple, the adjective *lilac* adds nothing to the overall interpretation and this violates the NVP. In

addition, we need something due to which the adjective *lilac* overrides the information about the color of the peel, rather than the color of the stem. Intuitively, it is clear that some aspects of a concept are more central to it than others. This is also shown in experimental studies where people have to rank properties of a concept as more or less related to it or in the speed of the response when people are asked to verify whether a property is true of a particular concept (Glass and Holyak 1975 in Smith, Osherson, Rips and Keane 1988). Smith, Osherson, Rips and Keane (1988) propose a model of conceptual representation in which the prototype of a concept consists of attribute value pairs in combination with a value indicating the salience of a value and information about the diagnosticity of an attribute, which is a measure of how important the attribute is in discriminating instances of the concept from instances of other concepts. For example, Smith et al. list three attributes of the concept apple: color, shape and texture, to which they respectively assign the values 1, 0.5 and 0.25. Each attribute is specified for three possible values with a number indicating the salience of each value (color: red 25, green 5, brown 0, shape: round 15, square 0, cylindrical 5, texture: smooth 30, rough 5, bumpy 0). Smith et al. suspect that the salience of a value depends on the subjective frequency with which the value occurs and the perceptibility of the value. Knowledge like this can be incorporated by treating the constraint FAITH not as a general constraint that is violated as many times as there are features missing in a candidate with respect to the input features, but as a family of ranked constraints that express faithfulness to a particular feature. For every feature in the input, there is a faithfulness constraint that is ranked with respect to other faithfulness constraints, pertaining to other features. Now, as I mentioned before, this is a lot of information and spelling this all out may not seem that insightful. However, I think that nobody would deny that people possess knowledge like this (the abovementioned studies also show this is the case) so if it can be shown that this knowledge adds to compositional interpretation and if it can explain phenomena for which otherwise extra semantic knowledge has to be assumed, I think including this kind of knowledge can be insightful. The proposal to include world knowledge to account for interpretation is in line with the abovementioned work by Blutner (2004) in the sense that it tries to capture the influence of salience on interpretation. However, a crucial difference is that the approach by Blutner is non-compositional. Blutner argues that it is not possible to decompose the lexical items of a combined expression into conceptual components which combined together determine the conceptual interpretation of the whole expression. I argue that this is possible (at least to a certain extent). Note, however, that it is not compositional in the sense that words in isolation have a meaning and that from these meanings the meaning of a compound is built (as is argued to be necessary by Jansen 1997) but in the sense that the meaning of a compound is predictable based

on the meaning of its parts (which is compatible with the notion of functional compositionality as proposed by van Gelder 1990 or weak compositionality as proposed by Smolensky 1991). The relation between this type of compositionality and (bidirectional) OT is also addressed in de Hoop, Hendriks and Blutner (2007).

In line with the model by Smith, Osherson, Rips and Keane 1988, I propose that both the attributes and the possible values are ranked with respect to each other. For the current purpose I include only the attributes of the color of the peel and the attribute of the color of the stem (I could add, for example, the color of the pulp but the principle remains the same) and their possible values. Logically, the possible values of an attribute are ranked lower than the attribute itself. Also, I assume that the color of the stem of an apple is less salient than the color of the peel of an apple. If we assume that an apple is more likely to be red than to be green and that the stem of an apple is always brown, we get the following ranking of faithfulness constraints:

Faith color of(x, peel) >> Faith color of(red, peel) >> Faith color of(green, peel) >> Faith color of(x, stem) >> Faith color of (brown, stem)

Now we need a constraint that tells the interpreter to attach the adjective to the highest ranked attribute. I think this can be naturally done by extending the constraint NVP. If the adjective modifies a very low ranked attribute, say for example the color of a seed of an apple, the adjective is less significant to the overall interpretation than when it modifies a highly ranked attribute. So, I propose to reformulate the constraint as:

NVH: assign the most salient interpretation to a word (violated by candidates in which a word in the input doesn't contribute to the overall interpretation and candidates violate it for each higher ranked attribute that the adjective doesn't modify).

Let us now look at a tableau illustrating the interpretation of apple:

Input: $\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red} \vee \text{green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$	FIT	NVH	*NB	FAITH color of (peel, x)	FAITH color of (peel, red)	FAITH color of (peel, green)	FAITH color of (peel, yellow)	FAITH color of (stem, x)	FAITH color of (stem, brown).
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$						*	*		
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$					*		*		
$\lambda x[\text{apple}(x) \wedge \text{peel of (peel, x)} \wedge \text{color of (red} \vee \text{green, peel)} \wedge \text{stem of (stem, x)} \wedge \text{color of (brown, stem)} \wedge \text{etc.}]$				*					

Tableau 7: interpretation of *apple*

If the constraint pertaining to the attribute ‘color of peel’ is only satisfied by candidates that have a particular value for it, the candidate with the highest ranked value for it, ‘red’, wins. Note that while without further context typicality may play a role in interpretation, it is not the case that words are represented as prototypes. Instead, as I argued above, the meaning of words are overrepresented, including non-typical and possibly conflicting features (such as the possibility for an apple to be red or green). Also, words have a set of different but related interpretations. This means that in interpretation, prototypicality may play a role when there is no more specific information available but this does not mean that a less prototypical concept is less well described by a word. This means that there is no reason why a combination of *words* should lead to a combination of *prototypes* of the set of concepts the word can denote. Therefore, combining concepts defined in this way does not lead to the problem of compositionality discussed in for example Kamp and Partee (1995) and Osherson and Smith (1981) since concepts are more precise than words. For example, the adjective *red* can refer to a range of different colors but if someone has the concept that in English is best described as *red house*, the precise color and the part of the house that the adjective modifies are not undetermined. In contrast, a hearer

hearing the combination of words *red house* or *red apple* must solve this issue. This is illustrated for *lilac apple* in the following tableau:

Input: $\lambda x[\text{apple}(x) \wedge \text{peel of (peel, } x) \wedge \text{color of (red } \vee \text{ green, peel) } \wedge \text{stem of (stem, } x) \wedge \text{color of (brown, stem) } \wedge \text{etc.}]$ $\lambda y[\text{lilac}(y)]$	FIT	*NB	NVH	FAITH color of (peel, x)	FAITH color of (peel, red)	FAITH color of (peel, green)	FAITH color of (stem, x)	FAITH color of (stem, brown)
y = apple (peel = red, stem = brown)	**	*				*		
\hookrightarrow Y = peel					*	*		
Y = stem (peel = red)			*			*		*
Y = peel and peel = red	*					*		
peel = red			*			*		

Tableau 8: interpretation of lilac apple

The first candidate violates the constraint against non-selective binding. The third candidate violates the NVH constraint because there is a higher ranked appropriate attribute. The fourth candidate violates FIT because it has two conflicting values for the attribute color of peel. The fifth candidate violates NVH because the meaning of the adjective *lilac* is absent from the interpretation. Hence, the second candidate is the winner and the phrase *lilac apple* is interpreted as referring to an apple with a lilac peel.

4.3 OT lexical semantics and various types of adjectives

As I mentioned in section 4.1, Kamp and Partee (1995) show that a theory of lexical semantics that assumes that words are defined by a set of features can only explain the interpretation of adjective noun combinations when it concerns subsecutive adjectives. However, the problem that a feature based approach can only account for intersective adjectives only holds when the features are assumed to be the set of necessary and sufficient features that define the meaning of the word. If we allow stored features for a word to be absent in the winning candidate interpretation, privative adjectives can be modeled as well. Furthermore, the introduction of attribute value structures in combination with the mechanism of selective binding can account for subsecutive adjectives.

4.3.1 Intersective and subsective adjectives

Some examples of intersective adjectives given by Partee (2007) are *sick*, *carnivorous*, *blond*, *rectangular*, and *French*. Examples of subsective adjectives are *recent*, *good*, *perfect*. The subsective adjectives usually modify an activity in the typical examples they occur in, e.g. a good basketball player is good at playing basketball, a good surgeon is good at performing surgery. Pustejovski (1995) argues that *good* functions as an event predicate and modifies an event in a quale of the noun. For example a good knife modifies the telic role 'cut' in the qualia structure of the noun. The meaning of the noun *knife* can (partially) be represented as follows $\lambda x[\dots \text{Telic} = \lambda e \lambda y[\text{cut}(e, x, y)]]$. The adjective *good* modifies the event cut in the telic quale of the representation (Pustejovski 1995, p. 129). So here again the mechanism of selective binding plays an important role. Nominalized verbs can be represented in frames as follows (Loebner 2010):

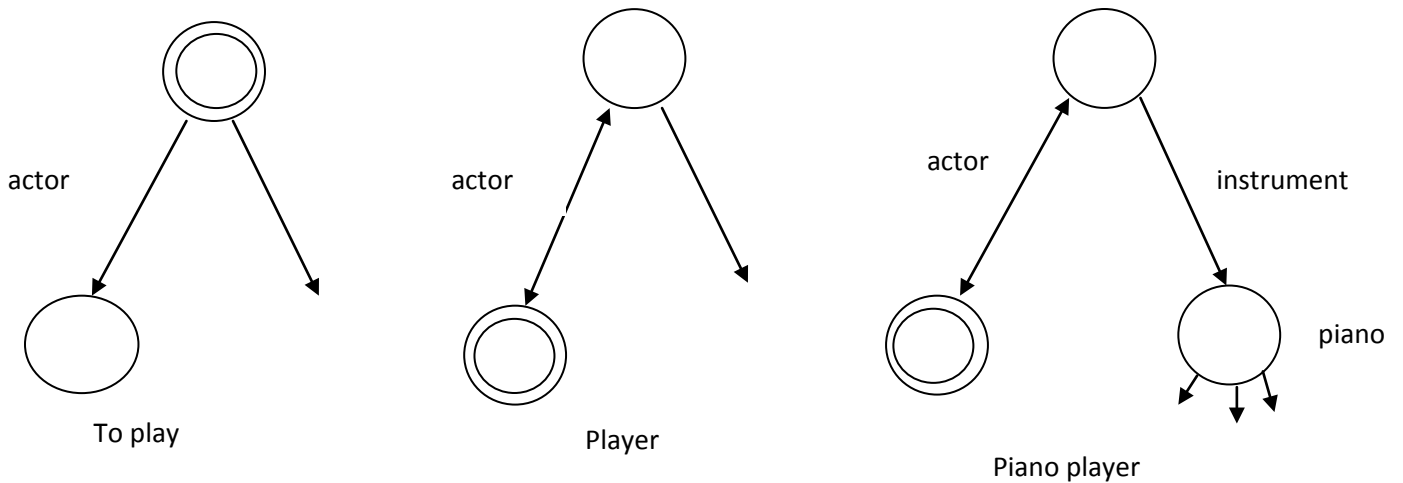


Figure 3: nominalized verbs

This corresponds to $\lambda x \exists y \exists e[\text{person}(x) \wedge \text{piano}(y) \wedge \text{play}(e, y, x)]$. Adding the adjective *good* leads to $\lambda x \exists y \exists e[\text{piano}(y) \wedge \text{play}(e, x, y) \wedge \text{good}(e)]$. The subsective adjectives are thus necessarily selectively bound if the noun denotes a person or object while the adjective modifies an event. This embedding of the adjective x in the representation of a noun A leads to a subsective interpretation because the event e_a that is part of the representation of A may be different from the event e_b in the representation of a noun B in a phrase xB , even though A and B have the same referent. It may also be the same, a good guitarist is also a good musician. The intersective adjectives on the other hand express properties that usually modify the head node in the frame. The value *carnivorous* (of the attribute say 'diet') in *carnivorous animal* for example pertains to the eating style of the whole animal, not just a part of it. The adjective is thus not selectively bound. This violates the constraint *NB but since the constraint NVH is

ranked higher this interpretation is still optimal. Since an intersective adjective y (typically) modifies the complete concept expressed by a noun A and by a noun B , and if A and B have the same referent, then if yA is true, yB is also true. I inserted *typically* in the previous sentence because I think there are cases where intersective adjectives are selectively bound, for example in *French restaurant*, where *French* pertains to the type of food that is served rather than the nationality of the building or organization. A French restaurant is therefore also not necessarily a French enterprise.

The analysis of subsective and intersective adjectives based on a difference between selective and non-selective binding illustrates my point about the contribution of conceptual knowledge to compositional interpretation. Partee (2001) proposes that the property of being a subsective or intersective adjective is stored in the lexicon as a meaning postulate. However, I think this property follows from the meaning of the adjective in combination with the noun and as such does not have to be stored independently.

4.3.2 Privative adjectives

Partee (2001) argues that privative adjectives like *fake* and *counterfeit* should be reanalyzed as subsective adjectives that trigger coercion of the nouns they modify. Such coercion can be motivated by treating the constraints on possible adjective meanings as presuppositions that must be satisfied by any use of an adjective. The corresponding coercion may then be seen as a kind of presupposition accommodation. Partee (2001) thus argues that there are no privative adjectives and that “normal adjectives” (excluding modal adjectives such as *alleged* or *possible*) are all subsective. Based on Polish NP-split phenomena and on sentences like *I don't care whether that fur is real fur or fake fur*, Partee argues that adjectives like *fake* are not privative but subsective and that the denotation of *fur* in such sentences is expanded to include real fur and fake fur. Indeed, the privative adjectives can be handled the same way as subsective adjectives, with the difference that due to conflicting features, the noun may lose more features than it does in combination with intersective or subsective adjectives. We already saw that combining color-denoting adjectives with nouns may lead to a cancellation of features expressed by the noun (e.g. red \vee green).

As a first example of an adjective often used to exemplify the class of privatives, consider the adjective *former*. Pustejovski (1995) analyzes the phrase *old friend* (as in ‘a friend for a long time’) as follows, where *old* modifies the event state that fulfills the telic role of the noun *friend*:

$$\lambda x \exists y [\text{formal} = \text{friend}(x, y) \wedge \text{Telic} = \lambda e^s [\text{friend_state}(e^s, x, y) \wedge \text{long}(e^s)] \dots]$$

The adjective *former* in for example *former wife* can be analyzed in a similar fashion. Let us say that, similar to the representation of *piano player*, the noun *wife* is defined as a person being in a state of being married. The adjective *former* indicates that this state no longer exists: $\lambda x \exists y \exists e^s [\text{person}(x) \wedge \text{female}(x) \wedge \text{married}(e^s, y, x) \wedge \text{ended}(e^s)]$.

The adjective *fake* or *counterfeit* would probably involve more changes to the modified noun, they negate a central property of the noun they modify (such as being made of animal skin for *fur*). In principle, this is compatible with OT lexical semantics because conflicting features may be omitted from a representation. However, further investigation is needed to find out how the adjective *fake* systematically negates the appropriate feature for several nouns.

4.3.3 Stone Lion

An adjective noun combination that is also said to involve (metonymic) coercion is *stone lion*. The noun *lion* is coerced from referring to the whole animal to only an image of the animal. This has always been a problematic example because here the adjective seems to change the denotation of the noun, which is the head. Partee (2001) already hinted at an OT solution for phrases like this involving competition between the non vacuity principle and the head primacy principle.

I mentioned that FIT is violated in three cases: 1: when there are two or more values for the same attribute; 2: when a value or attribute is of the wrong type as determined by a type hierarchy, for example the value *warm* for the attribute *color* or the attribute *color* for the type *smell*, and 3: when a value of an attribute is in incongruence with an inherited value. In a type hierarchy, subtypes inherit all the information from their supertypes. Information in the supertype cannot be overridden by applying selective modification in a subtype because the information is only inherited downwards in the hierarchy.

Now suppose we have a type hierarchy including the following sub hierarchy: *animate entity* >> *mammal* >> *lion*. Since animate entities are by definition of organic material, this is specified at the level of ‘animate entity’ and inherited by ‘mammal’ and ‘lion’. Now say the representation of lion is as follows, including inherited features.

$\lambda x [\text{Lion}_0(x) \wedge \text{animate}_2(x) \wedge \text{material of}_2(\text{organic}, x) \wedge \text{mamal}_1(x) \wedge \text{suckles young}_1(x) \wedge \text{color of}_0(\text{yellowish}, x) \wedge \text{image of}_0(\text{lion image}, x) \text{ etc.}]$

The attributes are marked for whether they are inherited or not and from what level. The adjective only modifies the type denoted by the noun and cannot make changes inside the higher levels. However, it

can cause, for example, a cancellation of the feature *animate* and all information that comes with it since this would not lead to inconsistencies in the representation of *lion*. This can be seen in the following tableau.

λx [$Lion_0(x) \wedge animate_2(x) \wedge material\ of_2(organic, x) \wedge mammal_1(x) \wedge suckles\ young_1(x) \wedge color\ of_0(yellowish, x) \wedge image\ of_0(lion\ image, x).$] $\lambda x[stone(x)]$	FIT	NVH	*NB	FAITH
λx [$Lion_0(x) \wedge animate_2(x) \wedge material\ of_2(\mathbf{organic}, x) \wedge mammal_1(x) \wedge suckles\ young_1(x) \wedge color\ of_0(yellowish, x) \wedge image\ of_0(lion\ image, x) \wedge \mathbf{stone}_0(x)$]	*		*	
λx [$Lion_0(x) \wedge animate_2(x) \wedge material\ of_2(organic, x) \wedge mammal_1(x) \wedge suckles\ young_1(x) \wedge color\ of_0(yellowish, x) \wedge image\ of_0(lion\ image, x)$] (stone is not interpreted)		*		
$\Rightarrow \lambda x$ [$Lion_0(x) \wedge non-animate(x) \wedge color\ of_0(yellowish, x) \wedge image\ of_0(lion\ image, x) \wedge stone_0(x)$]			*	****

Tableau 9: interpretation of *stone lion*

Since the adjective only modifies the level at which the noun denotes (not its supertype), candidates in which the adjective overrides information at level one or two are not included. Now, if the adjective *stone* modifies *lion* with the properties ‘animate’ and ‘mammal’ intact, as in the first candidate, this leads to a violation of FIT because lion inherits the value ‘organic’ for the attribute ‘material’. Another option is to not interpret *stone*, as in the second candidate, which leads to a violation of NVH. In the third candidate, the levels one and two are removed from the representation of *lion* due to which *stone* is not in conflict with the value ‘organic’ anymore. This means that this particular representation expressed by the noun *lion*, is not placed in the hierarchy as a subtype of animal and mammal, due to which it does not inherit the features specified for those types and due to which it is not interpreted as an instance of an animate entity.

6. Conclusions

I had two goals in writing this paper. The first goal was to explore the consequences of an OT approach to lexical semantics for the view on the relation between form and meaning. I argued that there are three major consequences. The first is that there is no strict relation between form and meaning. A meaning is not a static property of a form but it is the input to the process of production or it is the result of the process of interpretation. Consequently, the meaning of a word only 'exists' when it is being used. The second consequence is that word meanings are overspecified. This follows from the choice for OT but I also gave several independent arguments for this position. By means of Zwarts' (2004) analysis of *(a)round*, I illustrated the third consequence, which is that competition is important.

My second goal was to explore the usefulness of OT for the analysis of a broader set of lexical items, in particular content words. The set of analyses I discussed here is too limited to give a conclusive answer to this question. Nonetheless, some things can be said. First of all, the analyses of content words seem to be more complex than the analyses of function words. To be able to account for the interpretation of the noun *apple*, extra constraints were needed as well as a more structured representation of semantic features. A second thing to note is that the analysis in this paper borrows a lot of insights from the Generative Lexicon (Pustejovsky 1995). This is not surprising since GL also assumes word senses with a greater internal structure than is usually assumed. In OT lexical semantics this view is extended to the point of overspecification. A possible benefit of OT lexical semantics over GL is that it can account for additional phenomena such as color-denoting adjectives and the interpretation of adjective-noun pairs such as *stone lion*. Future work in OT lexical semantics should demonstrate whether it can account for other phenomena such as the semantics of verbs, other forms of type coercion and dot objects like *book* which can refer to the physical object (as in *a red book*) and to the non-physical content (as in *an interesting book*).

Another benefit of OT lexical semantics could be that no extra semantic knowledge has to be assumed. In GL a lot of information is represented at the semantic level, which is additional to a conceptual level. The analysis in this paper showed that conceptual knowledge in some domains behaves systematically and is able to explain phenomena for which otherwise extra semantic knowledge has to be assumed. People have detailed knowledge about concepts and intuitions about which aspects are more central to the concept than others. If it can be shown that this knowledge leads to compositional interpretation in other domains as well and no extra knowledge has to be assumed, overspecification may eventually be the most economical way of organizing the lexicon.

A last positive aspect of OT lexical semantics is that it analyses the interpretation of words with similar mechanisms as those that have been used to analyze other linguistic domains. Optimization has proven very successful in phonology and (perhaps to a lesser extent) in syntax, semantics and pragmatics. Further development of OT lexical semantics would therefore contribute to the development of a grammatical framework that pertains to all types of linguistic knowledge.

References

- Barsalou, L.W. (1992). Frames, concepts, and conceptual fields. In: E. Kittay & A. Lehrer (eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization*. Hillsdale, NJ: Erlbaum.
- Bierwisch, M. (1983). Semantische und konzeptuelle Repräsentation lexikalischer Einheiten. In: W. Motsch & R. Ruzicka (Hrsg.), *Untersuchungen zur Semantik*, Akademie Verlag, Berlin.
- Bierwisch, M. & Schreuder, R. (1992). From Concepts to Lexical Items. *Cognition* 42, 23-60.
- Blutner, R. (2004). Pragmatics and the lexicon. In: L. R. Horn & G. Ward (eds.) *Handbook of Pragmatics*, Oxford, Blackwell, 2004
- Blutner, R. (1998). Lexical underspecification and pragmatics. In: P. Ludewig & B. Geurts (eds.), *Lexikalische Semantik aus kognitiver Sicht*. Gunter Narr Verlag, Tübingen.
- Blutner, R., Hendriks, P., Hoop, H. de, & Schwartz, O. (2004). When compositionality fails to predict systematicity. In: S. Levy & R. Gayler (eds.), *Compositional Connectionism in Cognitive Science*. Papers from the AAAI Fall Symposium. Technical Report FS-04-03.
- Boersma, P. (1998). Functional Phonology. PhD. Amsterdam.
- Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity, *Linguistics and Philosophy* 21, 159–210.
- Fodor, J., & McLaughlin, B. (1991). Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work. In: T. Horgan & J. Tienson (eds.), *Connectionism and the Philosophy of Mind*. Cambridge, MIT press/Bradford books.
- Fodor, J & Lepore, E. (2002). *The compositionality papers*. Oxford, Clarendon Press.
- Fodor, J. A. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis, *Cognition* 28, 3-71.
- Fong, V. (2005). Unmarked *already*: Aspectual expressions in two varieties of English. In: H.j. Verkuyl, H. de Swart & A. van Hout (eds.), *Perspectives on Aspect*. Dordrecht, Springer.
- Gelder, T. van (1990). Compositionality: A Connectionist Variation on a Classical Theme, *Cognitive Science* 14:355-84
- Glass, A. & Holyoak, K. (1975). Alternative conceptions of semantic memory, *Cognition*, 3, 313-339.

Hendriks, P. & Hoop, H. de (2001). Optimality Theoretic Semantics. *Linguistics and Philosophy* 24, 1-32.

Hogeweg, L. (2009a). The meaning and interpretation of the Dutch particle *wel*, *Journal of Pragmatics*, 41:3, 519-539.

Hogeweg, L. (2009b). *Word in process. On the interpretation, acquisition and production of words*. Doctoral dissertation, Radboud University Nijmegen.

Hoop, H. de, Hendriks, P. & Blutner, R. (2007). On compositionality and bidirectional optimization. *Journal of Cognitive Science* 8, 137-151.

Hoop, H. de & Swart, H. de (2000). Temporal adjunct clauses in Optimality Theory. *Rivista di Linguistica* 12.1, 107-127.

Jackendoff, R. (1996). Conceptual Semantics and Cognitive Semantics, *Cognitive Linguistics* 7, 93-129.

Jackendoff, R. (in press), Conceptual Semantics, in *Semantics: An International Handbook of Natural Language Meaning*.

Janssen, T. (1997). Compositionality. In J.van Benthem and A. ter Meulen (eds), *Handbook of Logic and Linguistics*, 417–74. Elsevier/MIT Press, Amsterdam/Cambridge Mass.

Kamp, H. (1975). Two theories about adjectives. In E.L. Keenan (ed.), *Formal Semantics for Natural Language*, Cambridge University Press, Cambridge, 123-155.

Kamp, H. & Partee, B. (1995). Prototype theory and compositionality, *Cognition* 57, 129-191.

Katz, J. and Fodor, J. (1963). The structure of a semantic theory, *Language* 39: 170–210.

Keenan, E.L. (1974), The functional principle: Generalizing the notion of Subject of. In: *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, Chicago, Illinois, 298-310.

Lahav, R. (1989). Against compositionality: the case of adjectives, *Philosophical studies*, 57.

Legendre, G., J. Grimshaw, & S. Vikner (eds.). 2001. *Optimality-theoretic Syntax*. MIT Press.

Montague, R. (1970). Universal Grammar, *Theoria*, 36, 373-398.

Loebner, S. (2010) Frame Semantics as Semantics. Paper presented at the workshop *Frame Semantics as Semantics*, Amsterdam, January 22-23, 2010.

Osherson, D. N. & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts, *Cognition*, 9, 35-58.

Partee, B. (2001). Privative Adjectives: Subjective plus Coercion. In: T.E. Zimmerman (Ed.), *Studies in Presupposition*.

- Partee, B. H. (2007): Compositionality and coercion in semantics: The dynamics of adjective meaning. In: G. Bouma et al. (eds.): *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 145-161.
- Petersen, Wiebke (2007). Representation of Concepts as Frames. In: *The Baltic International Yearbook of Cognition, Logic and Communication*, Vol. 2, p. 151-170
- Petersen, W., Fleischhauer, J., Beseoglu, H., Bücken, P. (2008). A frame-based analysis of synaesthetic metaphors. In: *The Baltic International Yearbook of Cognition, Logic and Communication*, Vol. 3, New Prairie Press.
- Prince, A. and Smolensky, P (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder. Revised version published by Blackwell, 2004.
- Pustejovsky J. (1995). *The Generative Lexicon*. Cambridge, MIT Press.
- Quine, W.V. (1960). *Word and Object*. Cambridge, MIT Press.
- Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation and deduction, *Journal of Semantics*, 10:123-179, 1993.
- Rooy, R. van (2003). Negative Polarity Items in Questions: Strength as Relevance, *Journal of Semantics* 20, 239-273.
- Rumelhart, D., Smolensky, P. McClelland, J., & Hinton, G. (1986). Schemata and sequential thought processes in parallel distributed processing. In: *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2, Psychological and biological models*, D. E. Rumelhart, J. L. McClelland and the PDP Research Group, 7-57. MIT Press.
- Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model, *Cognitive Science*, 12, 485-527.
- Smolensky, P. (1991). Connectionism, constituency and the language of thought. In B. M. Loewer & G. Rey, (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford, Blackwell.
- Smolensky, P. and Legendre, G. (2006). *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar Vol. 1: Cognitive Architecture; vol. 2: Linguistic and Philosophical Implications*. Cambridge, MIT Press.
- Soderstrom, M. Mathis, D and Smolensky, P. (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory. In: P. Smolensky & G. Legendre (eds.) *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar. Vol 2: Linguistic and Philosophical Implications*. Cambridge, MIT Press.
- Winter, Y. (2001). Plural predication and the Strongest Meaning Hypothesis. *Journal of Semantics* 18, 333-365.
- Winter, Y. (2010), *Between Logic and Common Sense: The Formal Semantics of Words*. Vidi-application.

Zeevat, H. (2001). The asymmetry of optimality theoretic syntax and semantics, *Journal of Semantics*, 17(3):243-262, 2001.

Zeevat, H. (2002). Explaining presupposition triggers. In K. van Deemter & R. Kibble (eds.), *Information Sharing*, 61-87. Stanford, CSLI Publications.

Zeevat, H. (2006). Discourse structure in optimality theoretic pragmatics. In C. Sidner, J. Harpur, A. Benz & P. Kühnlein (eds.), *Proceedings of the Workshop on Constraints in Discourse*, 155–161.

Zwarts, J. (2004). Competition between Word Meanings: The Polysemy of (A)Round. In Meier, C. and Weisgerber, M. (eds.), *Proceedings of SuB8*. Konstanz, University of Konstanz Linguistics Working Papers.

Zwarts, J. (2008). Priorities in the production of prepositions. In A. Asbury, J. Dotlacil, B. Gehrke & R. Nouwen (eds.), *Syntax and Semantics of Spatial P*. Amsterdam, John Benjamins.