Negating Gradable Adjectives

Abstract: In this paper, I give an analysis of the syntax and semantics of the prefix *un*- with gradable adjectives like *unhappy*, *unfriendly*, *unsafe*, *uninteresting*, and compare it to the syntax and semantics of *not*. Within the framework of Collins and Postal 2014, I propose that *un*- and *not* have the same semantics but negate different constituents, accounting for differences in interpretation.

Key Words: gradable adjectives, negation, degree quantification, quantifier domain restriction, litotes, Klima tests

1. Introduction

Consider the following two sentences:

- (1) a. John is unhappy.
 - b. John is not happy.

In both cases there is a negative morpheme, *un*- in (1a) and *not* in (1b). And furthermore, the two sentences overlap in truth conditions. It seems that (1a) entails (1b) but not vice a versa. Certainly, if John is unhappy, we can conclude that he is not happy. But if he is not happy, he may not be unhappy either (but somewhere in the middle of the scale of happiness which ranges from very unhappy to very happy).

I propose that both *un*- and *not* are negative morphemes of the category NEG (see Collins and Postal 2014). But in (1a) *un*- modifies the adjective while in (1b) *not* modifies a covert degree quantifier phrase. I show how the difference in interpretation between (1a) and (1b) follows from this structural assumption.

In section 2, I introduce the Scale of Happiness which is partitioned by the two predicates *happy* and *unhappy*. Section 3 addresses a compositionality issue that arises in introducing the external argument. Section 4 motivates a covert degree quantifier. Section 5 discusses the difference between *happy* and *unhappy*. Section 6 discusses the difference between *unhappy* and *not happy*. Section 7 shows how the truth conditions of litotes examples are calculated. Section 6 shows how the Klima tests provide support for my analysis. Section 8 is the conclusion.

This paper focusses on gradable adjectives like *happy*. Whether its conclusions extend to the use of *un*- with non-gradable adjectives is left for future research (on the range of adjectives taking *un*- prefixation see Horn 2001). Furthermore, for brevity sake, the paper focuses exclusively on the predicate uses of such adjectives (*John is unhappy*), even though attributive uses are also possible (*the unhappy person*). For more on the syntax of *un*-, see Kayne 2017 and De Clerq and Vanden Wyngaerd 2018.

There have been other approaches to *un*-prefixation in the literature, including Horn 2017, Krifka 2007 and Blutner 2004. These approaches all focus for the most part on pragmatics (e.g., the issue of pragmatic strengthening of a contradictory negation to a contrary, or the issue of the mitigating effect of double negation). The focus of my paper is rather on the semantic values of *un*- and *not*, the truth conditions of sentences involving *un*- and *not* and entailment relations between the various sentences. As far as I know, my analysis is the first attempt to give a compositional analysis of the semantics of sentences involving *un*-prefixation. Therefore, my approach is not directly comparable to these other approaches, because my goals are different.

However, where appropriate, I will discuss the data analyzed by these other approaches and the assumptions that they make.

2. Scale of Happiness

A standard way to define the semantic value of a gradable adjective is as follows (see Kennedy and McNally 2007: 349):

(2)
$$[[happy]] = \lambda d\lambda x.x$$
 is happy to degree d

This means that *happy* takes two arguments, a degree and an individual, and is true if the individual x is happy to degree d. I will need to modify this semantic value below in light of facts concerning *un*-modification.

Whether *un*- combines with an adjective or a whole adjective phrase is not relevant to the present paper:

(3) a. [NEG ADJ] b. [NEG ADJP]

Since *un*- is NEG, I assume that its semantics is given by the semantics of negation in Collins and Postal 2014:

(4) If X has a semantic type ending in $\langle t \rangle$, then NEG takes X with semantic value: $\lambda P_1 \dots \lambda P_n [\dots]$ And returns Y with semantic value: $\lambda P_1 \dots \lambda P_n \neg [\dots]$

For propositional variables p (no predicate abstraction), the negation is simply $\neg p$. On this view, negation can combine with constituents of various different types, parallel to the analysis of conjunction given in Partee and Rooth (1983). One way to think of it is that the different types partition the domain of the negation function.

Assuming the structure (3a) for convenience, and applying (4), the semantics of *un*- are given below:

(5) a. $\llbracket un- \rrbracket = \lambda X.\lambda d.\lambda x.\neg X(d)(x)$ b. $\llbracket un-happy \rrbracket = \lambda d\lambda x.\neg happy(d)(x)$

The problem with (5b) is that it makes the claim that x is unhappy to degree d if it is not the case that x is happy to degree d, and hence does not distinguish (1a) and (1b). In other words, given (5), the phrase *John is unhappy* would not actually entail that John is unhappy (but only not happy to a certain degree).

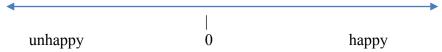
I propose instead that *happy* should be defined as a predicate of degrees. Then the semantic value of *unhappy* will be given in (6b):

(6)	a.	[[happy]]	=	$\lambda d.happy(d)$
	b.	[[un-happy]]	=	λd . \neg happy(d)

Defining *happy* and *unhappy* in this manner avoids the problem noted following (5). Suppose John's degree of happiness is d, then John is unhappy iff \neg happy(d), which accurately yields the truth conditions. Of course, it remains to see how the external argument is linked to particular degrees. I address that issue in the next section.

The range of degree values can be diagrammed as follows:

(7) Scale of Happiness



Anything to the right of 0 is happy and anything to the left of 0 (including 0) is unhappy. Therefore, *un*-prefixation yields a contradictory, not a contrary negation (for a similar assumption, see Krifka 2007: 170 and De Clercq and Vanden Wyngaerd 2018, for *un*-prefixation as contrary negation see Horn 2001: 275, 2017: 8).

For the analyses in this paper, I need to be more specific about what degrees are. Following Kennedy 2001: 52, I define a scale S to be "...a linearly ordered, infinite set of points, associated with a dimension that indicates the type of measurement that the scale represents."

Deviating from Kennedy (who does not discuss *un*-prefixation), I will define as a degree any subset of S of the following form [0, x] (x greater than or equal to 0 on the real number line) or [x, 0] (x less than or equal to zero on the real number line). Furthermore, ordering amongst the degrees is defined as follows:

(8) a. d1 > d2 iff $d1 \cap d2 = d2 \wedge d1 \neq d2$ b. d1 < d2 iff $d1 \cap d2 = d1 \wedge d1 \neq d2$

Concretely consider the following example:

According to the definitions in (8), d1 is greater than d2 and d4 is greater than d3. Degrees on the opposite sides of the scale are not ordered. For example, for d2 and d3, neither of the ordering relations in (8a,b) holds.

A clarification is in order. I am representing the scale of happiness on the real number line, with a 0 element. I chose this representation for three reasons. First, it allows me to define degrees on the happy and unhappy part of the scale and how they are ordered. Second, it provides a single set to which all the degrees belong. Third, it provides an intuitive graphic representation, which I will employ throughout the paper. But the orderings referred to in the rest of the paper are not the orderings of real numbers, but rather orderings of degrees as defined in (8).

This scale makes sense of comparative constructions involving *un*-prefixation. First, if both John and Bill are happy, one can say (10):

(10) John is happier than Bill

"The degree to which John is happy exceeds the degree to which Bill is happy."

But in the same situation, it is not possible to say (11) (# means that it is not felicitous):

(11) #Bill is unhappier than John."The degree to which Bill is unhappy exceeds the degree to which John is unhappy."

The reason for the infelicity of (11) is that neither Bill nor John is unhappy, so they are not associated with degrees in the left part of the happiness scale.

With this background, return to the semantic values of *happy* and *unhappy*. First, *happy* can be defined in terms of degrees as follows, where DEG_h is a predicate that is true of degrees on the happiness scale.

(12)	a.	[[happy]]	=	λd : DEG _h (d). happy(d)
	b.	[[un-]]	=	$\lambda P .\lambda d. \neg P(d)$
	c.	[[un-happy]]	=	λd : DEG _h (d). ¬happy(d)

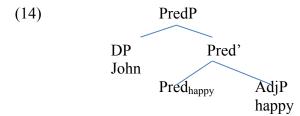
And the following condition holds: $\forall d[DEG_h(d) \rightarrow happy(d) \lor \neg happy(d)]$.

3. Adding the External Argument

Given the semantics in (12), *happy* is a predicate of a degree variable. So the question is how to incorporate an external argument (the experiencer of the emotion) as in sentences like (13):

(13) John is happy.

Following Bowers 1993 (see also Kratzer 1996 for related ideas), I propose that part of the structure of sentences like (13) is (leaving out the copula verb and TP):



Given this tree, I define the semantic value of Pred in (15). degree_h is a function which takes an individual and returns its position on the scale of happiness. I assume that the degree_h function is defined for all humans: Everybody is happy or unhappy to some degree. That is, everybody has unique position on the happiness scale (at a particular point in time, which is not represented in the semantic values):

(15)
$$\llbracket Pred_{happy} \rrbracket = \lambda P.\lambda x.\lambda d. degree_h(x) = d \land P(d)$$

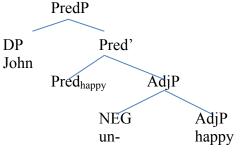
For example, the semantic value of the structure in (14) is calculated as follows:

(16)
$$\llbracket (14) \rrbracket = [\lambda P.\lambda x.\lambda d. degree_h(x) = d \land P(d)](\llbracket happy \rrbracket)(\llbracket John \rrbracket)$$

= $\lambda d. degree_h(John) = d \land happy(d)$

Consider now *unhappy* in the sentence *John is unhappy*. The tree below gives part of the structure of the sentence:

(17)



The semantic value of (17) is the following:

(18) $\llbracket (17) \rrbracket = \lambda d. degree_h(John) = d \land \neg happy(d)$

With these preliminaries out of the way, I return now to providing an analysis of the difference between (1a) and (1b).

4. Covert Degree Quantifier

Consider first (19a) which can be paraphrased as (19b):

- (19) a. John is happy.
 - b. John is happy to some degree.

I suggest that (19a) involves existential quantification over degrees, as in the overt degree phrase in (19b), so that (19a) has the syntactic structure in (20a) or (20b), where caps indicate non-pronunciation.

(20)	a.	John is happy [TO [SOME DEGREE]]
	b.	John is [SOME DEGREE] happy.

See (29) below for evidence supporting the order in (20b), where the covert degree quantifier phrase precedes the adjective.

Furthermore, I suggest that this existential quantification has a restricted domain. Normally, when one says (19a) one does not mean that John is happy to some small or insignificant degree, but rather he is happy to some significant extent. This range of degrees of happiness is seen in expressions such as the following:

- (21) a. John is a tiny bit happy.
 - b. John is sort of happy.
 - c. John is somewhat happy.
 - d. John is reasonably happy.

- e. John is happy.
- f. John is quite happy.
- g. John is very/really happy.
- h. John is really very happy.
- i. John is extremely happy.

Because of scalar implicatures, sentences like (21a-d) also imply that John is not any happier than the amount specified in the degree expression. Without any degree modification, and minimal context, (21e) falls in the middle of the range of possibilities. So I suggest that the existential quantifier in (20) is the following:

(22) some degree greater than or equal to a contextually given degree n_1

I use n_1 to distinguish it from n_2 which will be introduced below (when talking about the unhappy part of the scale of happiness). The existential quantifier and its quantifier domain restriction can be represented as follows:

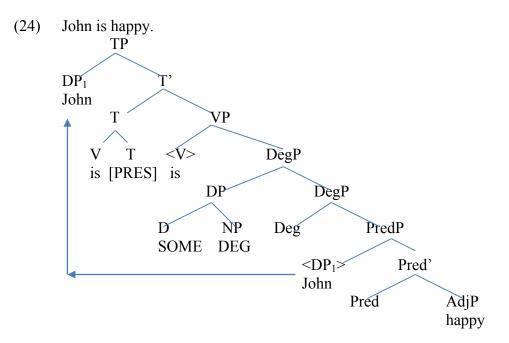
(23) [SOME DEG_h]] = $\lambda P[\exists d (P(d) \land DEG_h(d) \land d \ge n_1)]$

The DEG_h predicate is true or false of degrees falling on the scale of happiness (see (7)), and undefined otherwise. Since all quantification in this paper is over the degrees on the scale of happiness, I will henceforth leave DEG_h out of the sematic values.

An additional question, which I will not deal with in this paper, is how the domain restriction $(d \ge n_1)$ in (23) is syntactically represented. There is a large literature on this topic, and it is not relevant to the analysis in this paper (see Elbourne (forthcoming) for discussion and references).

In the literature on gradable adjectives, the function of (23) is attributed to POS, as in Kennedy and McNally (2005: 350): "...unmodified APs actually contain a null degree morpheme *pos* (for positive form) whose function is to relate the degree argument of the adjective to an appropriate standard of comparison...". From this point of view, (23) is not an innovation, but rather corresponds to what is assumed quite generally in the literature on the semantics of adjectives.

Putting the assumptions in (14) (about adjectives) and (23) (about degree quantification) together, a tree diagram for (19a) is as follows. I assume that Spec DegP is occupied by the degree expression (and that Deg itself denotes the identity function). As usual, V to T is obligatory with finite auxiliaries.



Given (24), the truth conditions of (19a) are as follows:

(25) John is happy

 $\begin{bmatrix} \lambda P. \exists d (P(d) \land d \ge n_1) \end{bmatrix} (\lambda d. degree_h(John) = d \land happy(d)) \\ = \exists d(degree_h(John) = d \land happy(d) \land d \ge n_1) \end{bmatrix}$

This can be paraphrased as follows: there is some degree d on the happiness scale, greater than or equal to a contextually given degree n_1 , such that John is happy to degree d. These are intuitively the right truth conditions.

Quantifier domain restriction is influenced by contextual factors. For example, consider the following exchange:

(26) a. Is John happy with the hiring decision?

b. I guess you could say that, but he is not overjoyed.

(26b) says that John is happy, but n_1 is being pushed down from its usual standard position by use of the phrase 'I guess' and by the continuation, 'but he is not overjoyed'. This is a completely different value of n_1 than if the sentence *John is happy with the hiring decision* were used out of the blue. So this example shows that there is flexibility in the value of n_1 .

Krifka 2007: 172 suggests that uncertainty might also play a role in determining the use of *happy* and *unhappy*: "As a consequence of this uncertainty about the location of the border between happiness and unhappiness, the use of *unhappy* and *happy* is pragmatically restricted to those areas for which the interlocutors can assume to be in mutual agreement, to ensure safe communication." In my theory, such pragmatic factors could play a role in setting n_1 . In thinking how to describe a situation, I choose a value of n_1 for which I know that my interlocutor will agree that John is happy (some robust level of happiness).

Consider now sentence (1b). Once again following Collins and Postal 2014, the negation of (19) is as follows.

(27) $[[\text{NEG SOME}] \text{ DEGREE}_h] = \lambda P \neg [\exists d (P(d) \land d \ge n_1)]$

Note that I am assuming (as in CP2014) that negation negates quantifiers directly. It would not change the analysis very much to assume a clausal NEGP. I leave it to the reader to show that the two analyses (clausal NEGP versus NEG modifying SOME) yield equivalent truth conditions.

Then (1b) has the following truth conditions (which is the negation of (25)):

(28)
$$\neg \exists d(degree_h(John) = d \land happy(d) \land d \ge n_1)$$

Some syntactic evidence for a negative degree quantifier phrase in examples like (1b) is provide by the following sentence:

(29) John is not at all happy.

Note that *at all* usually modifies negative DP such as the following:

- (30) a. Nobody at all was there.
 - b. *Every person at all showed up.
 - c. *Some people at all showed up.

At all has the effect of strengthening the quantification by lifting domain restrictions. So the claim is that (29) involves a negative existential degree quantifier modified by at all, which has the effect of lowering n_1 to 0 (see the diagram in (34)).

Another piece of evidence for a degree quantifier phrase in examples like (1b) is that the degree quantifier phrase sometimes appears overtly:

(31) a. John is not a bit happy.

b. John is not the least bit happy.

In these examples, the expressions [not a bit] and [not the least bit] seem to be overt occurrences of the degree quantifier phrase postulated in (22).

5. *happy* versus *unhappy*

Consider adding an existential degree quantifier to examples with *unhappy*. Since John is on the unhappy portion of the scale of happiness, I assume that n_2 (the standard value) will also be on that portion of the scale (guaranteeing that $d \ge n_2$ is defined). Since unhappy(d) holds for d = 0, it follows that n_2 can be zero as well. The resulting semantic value (derived compositionally) is given below:

(32) $\exists d(degree_h(John) = d \land \neg happy(d) \land d \ge n_2)$

Recall that $d \ge n_2$ means that d and n_2 are in the unhappy part of the scale in (9), and that d is to the left of n_2 (that is, further away from zero, see the definitions of inequality in (8)).

Now compare John is happy and John is unhappy in terms of these truth conditions:

(33) a. John is happy.

$$\exists d(degree_h(John) = d \land happy(d) \land d \ge n_1)$$

b. John is unhappy.
 $\exists d(degree_h(John) = d \land \neg happy(d) \land d \ge n_2)$

These truth conditions can be illustrated on the scale of happiness as follows:

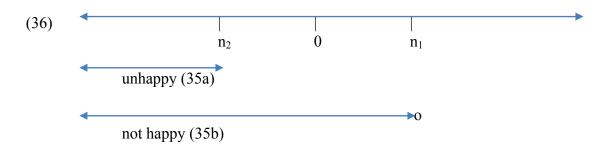
The first thing to note is that sentences with *happy* and *unhappy* come out as contraries, since the relevant sentences cannot be true at the same time, but they can be false at the same time (where John is neither happy nor unhappy). However, the fact that sentences with *happy* and *unhappy* are contraries is not stipulated on my account, but is rather derived from the presence of an existential quantifier with a domain restriction (as opposed to defining a special contrary negation, as in Horn 2017).

6. *unhappy* versus *not happy*

Now compare the two sentences in (1) and their semantic values, repeated below:

(35) a. John is unhappy. $\exists d(degree_h(John) = d \land \neg happy(d) \land d \ge n_2)$ b. John is not happy. $\neg \exists d(degree_h(John) = d \land happy(d) \land d \ge n_1)$

One can now see what accounts for the difference between (35a) and (35b). Consider the diagram in (306 (the o at the end of the second horizontal arrow represents not equal to):



The horizontal arrows under the scale in (36) are to be interpreted as follows: (35a) would be true if degree_h(John) were greater than or equal to n_2 (that is, to the left of n_2).

Similarly, (35b) would be true if degree_h(John) were a member of the unhappy part of the scale, or less than n_1 (that is, to the left of n_1) on the happy part of the scale.

From the diagram, it is clear that (35a) entails (35b), since the range of degrees of (35a) is a subset of the range of degrees of (35b). Furthermore, it is clear that (35b) does not entail (35a). It could be that John has some intermediate (not happy, not unhappy) state. The entailment relations are represented as follows:

- (37) a. John is unhappy \vDash John is not happy
 - b. John is not happy ⊭ John is unhappy

Consider now the following sentence from Krifka 2007: 167:

- (38) a. I am not happy at all, in fact, I am quite unhappy.
 - b. *I am quite unhappy, in fact, I am not happy at all.

The theory I propose has no problem accounting for this contrast. The intensifier *at all* in (38a) removes all domain restrictions (see the discussion following (30)) and has the effect of reducing n_1 to 0 (or very close to 0). It says that I do not occupy the happy part of the scale, therefore by implication I only occupy the unhappy part of the scale. The *quite* (which is a degree expression) before *unhappy* has the effect of pushing n_2 further to the left than the standard value (see (21f) for an analogous effect on the happy part of the scale). And so the second half of (38a) is a strengthening of the first half, which is consistent with the use of the phrase *in fact*. This also explains the unacceptability of (38b), since normally the stronger sentence needs to follow *in fact*, not precede it.

Horn 2017: 87 notes that contradictory negation is often strengthened to a contrary. For example, (39a) is can be continued with (39b). But spoken out of the blue, (39a) seems to imply (39c).

- (39) a. Alex isn't happy.
 - b. ...but he isn't unhappy either.
 - c. Alex is unhappy.

As noted above, (39c) entails (39a). So (39c) is stronger than (39a). As discussed by Horn and others, there are many ways a weaker statement can imply a stronger one (via R-implicature). My theory does not shed any new light on this issue, but it is consistent with the observations.

In general, manipulations of n_1 and n_2 in the domain restrictions are in part syntactic, semantic and pragmatic. So I am not claiming that my theory is a purely semantic or purely syntactic theory. Rather, it locates the nexus of these influences in the domain restrictions of the degree quantifiers.

7. Litotes: *not unhappy*

Horn 2017 cites the OED definition of litotes as 'a figure of speech in which an affirmative is expressed by the negative of a contrary.' A typical example of litotes is an expression like that in (40):

(40)John is not unhappy.

If both *not* and *un*- modified the adjective unhappy, then under the semantics of negation in Collins and Postal 2014, (40) should be equivalent to 'John is happy', but it is not. Rather, (40) is weaker than the positive (without double negation) (see Horn 2001: 306, Horn 2017: 89).

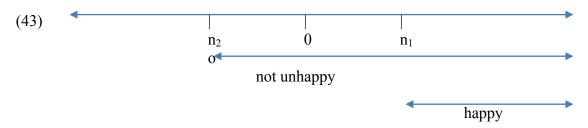
He was happy, or at least not unhappy. (41) a. b.

#He was not unhappy, or at least happy.

In these examples, the *at least* phrase introduces the weaker alternative. If she was happy, then she was not unhappy. But if she was not unhappy, it does not follow that she was happy. According to the theory developed so far, (40) has the following truth conditions:

(42) $\neg \exists d(degree_h(John) = d \land \neg happy(d) \land d \ge n_2)$

This can be diagrammed as follows:



This diagram accounts for the fact that John is happy entails that John is not unhappy, since the interval covered by *happy* is a strict subset of the interval covered by *not unhappy*. It is also clearly the case that John is not uhappy does not entail that John is happy for the same reason.

I represent these entailment relations as follows:

- (44) John is happy ⊨ John is not unhappy a.
 - John is not unhappy ⊭ John is happy b.

Horn 2017: 82 and Krifka 2007: 164 note that the double negation has a mitigating effect: "...double negatives have the same interpretation as weak positives." While my theory does not directly account for this observation, it is consistent with it. Even when double negation is interpreted as a weak positive, the entailment relations in (44) hold.

8. Klima Tests

Another fortunate consequence of the truth conditions in (35) is that they explain why (35a) does not count as sentential negation in the sense of Klima (1964). Consider the following generalization (from Collins and Postal 2017):

A sentence S is an instance of sentential negation only if some NEG or negative (45)quantifier DP takes widest scope in the matrix clause of S.

In (35a), the existential quantifier takes widest scope, and so (35a) does not count as sentential negation. In (35b), the negation takes widest scope, so (35b) counts as sentential negation. These predictions are correct, as shown below:

- (46) a. John is unhappy, isn't he?/I think.
 - b. John is not happy, is he?/I don't think.

In (46a), the negative tag-question and the positive parenthetical are used, as is expected when there is no sentential negation. In (46b), a positive tag-question and negative parenthetical are used, as is expected when there is sentential negation.

9. Conclusion

I have shown how it is possible to analyze *un*- and *not* as negative morphemes with the semantics of negation given in Collins and Postal 2014. The crucial difference is that *un*-modifies an adjective directly, whereas *not* modifies a covert degree quantifier (in the examples under consideration). I showed how my analysis explains entailment relations between various sentences and the Klima tests with *un*- and *not*.

References:

Blutner, Reinhard. 2004. Pragmatics and the Lexicon. In Laurence R. Horn and Gregory Ward (eds.), *The Handbook of Pragmatics*, 488-514.

Bowers, John. 1993. The Syntax of Predication. Linguistic Inquiry 24, 591-565.

Collins, Chris and Paul Postal. 2014. Classical NEG Raising. MIT Press, Cambridge.

Collins, Chris and Paul Postal. 2017. Interclausal NEG Raising and the Scope of Negation. *Glossa* 2.1, 1-29.

De Clercq, Karen and Guido Vanden Wyngaerd. 2018. Negation and the Functional Sequence. *Natural Language and Linguistic Theory*.

Elbourne, Paul. Forthcoming. Literal vs enriched meaning: It's raining. In Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann (eds.), *The Wiley Companion to Semantics*, Wiley.

Horn, Laurence. 2001. A Natural History of Negation. CSLI Publications.

Horn, Laurence. 2017. Lie-toe-tease: Double Negatives and Unexcluded Middles. *Philosophical Studies* 174, 79-103.

Kayne, Richard. 2017. Antisymmetry and Morphology. Prefixes vs. Suffixes. In Clemens Mayr and Edwin Wiliams (eds.) *Festschrift für Martin Prinzhorn, Wiener Linguistische Gazette*, 82, 145-161.

Kennedy, Christopher. 2001. Polar Opposition and the Ontonology of 'Degrees'. *Linguistics and Philosophy* 24, 33-70.

Kennedy, Christopher and Louise McNally. 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language* 81, 345-381.

Klima, Edward. 1964. Negation in English. In *The structure of language*. Jerry A. Fodor and Jerrold J. Katz (eds.) 246-323. Englewood Cliffs, New Jersey: Prentice Hall.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In J. Rooryck and L. Zaring (eds.), *Phrase Structure and the Lexicon*, 109-137. Dordrecht: Kluwer.

Krifka, Manfred. 2007. Negated Antonyms: Creating and Filling the Gap. In Uli Sauerland and Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan.

Partee, Barbara and Mats Rooth. 1983. Generalized Conjunction and Type Ambiguity. In Paul Portner and Barbara H. Partee (eds.), *Formal Semantics: The Essential Readings*. Blackwell.