RESEARCH

Asymmetric inference towards the antonym: Experiments into the polarity and morphology of negated adjectives

Nicolas Ruytenbeek[1], Steven Verheyen[2] and Benjamin Spector[2]

[1] Fonds de la Recherche Scientifique-FNRS, Research Centre in Linguistics (LaDisco), Université libre de Bruxelles (ULB), 50, av. F. R. Roosevelt, 1050, BE

[2] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL Research University, CNRS, Paris, FR

Corresponding author: Nicolas Ruytenbeek (nruytenb@ulb.ac.be)

In this paper, we investigate the interpretation of negated antonyms. A sentence such as *Peter is not tall* can be understood as meaning either that Peter is not tall *tout court* or that Peter is rather short (inference towards the antonym; ITA). We present the results of two experiments, in which we test two theoretical predictions. First, according to Krifka (2007), it is reasonable to expect a stronger ITA effect for positive versus negative adjectives. Second, elaborating on Krifka (2007), we expect ITA strength asymmetry to be greater for morphological antonymic pairs than for non-morphological pairs. In the first experiment, ITA strength was assessed implicitly by having participants judge the pragmatic acceptability of sentences involving a negated adjective and its antonym. In the second experiment, we collected explicit inferential judgments. The results of both experiments support the two predictions. We also discuss theoretical and methodological issues concerning the different notions of polarity.

Keywords: gradable adjectives; negation; antonymy; mitigation

1 Theoretical background

1.1 General introduction

Many adjectives have antonyms, that is they stand in opposition to adjectives with which they are incompatible (Lehrer & Lehrer 1982; Cruse 1986). For example, a man cannot be said to be both tall and short at the same time. According to the traditional view of the semantics of antonymic adjectives, items such as *tall* and *short* are *contraries* in that they allow for a middle term such as *medium* or *middle-sized* (Horn 1989: 6–21). For instance, the first part of a sentence such as (1) is perfectly coherent, but a sentence such as (2) is clearly a contradiction.

(1)   (Edith Templeton, *Gordon*, 1966)
        He was neither short nor tall, slender and narrow-boned, of an unimpressive physique I did not care for.

(2)   This integer is neither even nor odd.

Contrary adjectives stand in opposition to *contradictory* adjectives such as *odd* and *even*, or *dead* and *alive*, which follow the law of excluded middle. In the case of contradictory antonyms, the negation of one member of a pair should thus result in the affirmation of

the other member of that pair. But in the case of contrary antonyms, the negation of one member of a pair is not in general equivalent to the affirmation of the other member of that pair.

Contrary antonyms are generally *gradable, scalar* adjectives. For instance, both *tall* and *short* are associated with the scale of *height,* and are used to locate an object on this scale. An object is tall if its height is above a certain threshold θ_{tall}, and an object is short if its height is below a certain threshold θ_{short}, but an object whose height is between these two thresholds is neither tall nor short. This approach can be generalized to all pairs of contrary adjectives.

On this basis, the negation of an adjective such as *tall* corresponds to a scale region which properly contains the region corresponding to *short*, as illustrated in Figure 1, and so *not tall* does not entail *short.*

Things are, however, more complex. First, the threshold for *tallness* (θ_{tall}) is highly context-dependent. Second, speakers do not have in mind, in most contexts, a precise threshold for *tallness*. If someone's height is slightly above average, it is not clear whether the person is or is not tall—such cases are called *borderline* cases. In practice, when it is asserted that a person is tall, what is meant is that this person is *clearly* tall, i.e., *tall* is not normally used for someone whose height may or may not be above the threshold, depending on its precise value. To account for this fact, several approaches have been developed. In *trivalent approaches* to vagueness (see, e.g., Fine 1975), the view is that the sentence *Peter is tall* is *undefined,* rather than true or false, if Peter's height falls into the "unclear" zone, and its negation, *Peter is not tall*, comes out undefined as well. As a result, the relationship between *tall* and *not tall* is quite similar to that of *tall* and *short*: it is possible for a person to be neither clearly tall nor clearly not tall. In the *epistemic approach* to vagueness (Williamson 1994), the view is that the sentence *John is tall* is always either true or false, but that speakers do not have a complete grasp of its truth-conditions: there is a zone of heights such that speakers do not know whether a person whose height falls in this zone is tall or not tall. As a result, to be able to assert *Peter is tall* while respecting Grice's (1975) maxim of quantity, it has to be the case that Peter's height is located above this zone, and to be able to assert *Peter isn't tall*, Peter's height must be below this zone. In this approach as well, then, from the point of view of actual use, the relationship between *tall* and *not tall* becomes similar to that of *tall* and *short*: for both the pair *tall/short* and the pair *tall/not tall*, there is a zone of indifference such that neither member of the pair can be used to describe an individual whose height falls in this zone (see also von Stechow 2009).

Given that the boundaries of the zone of indifference are themselves highly context-dependent and vague, the conditions under which *not tall* is used might well be more or less the same as those that regulate the use of *short*. And in fact, there are contexts where *not tall* seems to imply *short*, or, similarly, *not large* will imply *small* and *not good* will imply *bad*. Consider the following examples:

(3)   a.   Rooms not large, but not small either.[1]
       b.   My bedroom is not large. Because I don't like small rooms, I spend little time in it.
       c.   Macaroni salad is just ok. Not good but not bad either.[2]
       d.   It's not good. Don't use it. (COCA Davies 2008)

---

[1] https://www.tripadvisor.com/ShowUserReviews-g293923-d580062-r13666773-Ha_Long_Bay_Hotel-Halong_Bay_Quang_Ninh_Province.html.
[2] https://www.yelp.com/biz_photos/uncles-shack-and-grill-brooklyn-2?select = M4Suv-oztO578SbG-32C9Ow.
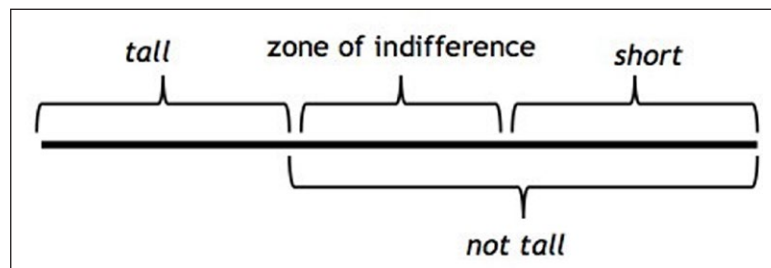
Figure 1: Ranges of application for *tall*, its negation, and its antonym *short.*

Somewhat impressionistically, while in (3a) and (3c), it has to be the case, for the discourse to be consistent, that *is not large* (resp. *is not good*) does not imply *small* (resp. *bad*), in (3b) and (3d), replacing *is not large* (resp. *is not good*) with *is small* (resp. *is bad*) does not seem to produce a significant change in meaning.

Whenever the negation of a scalar adjective licenses an inference to its contrary, we will describe this phenomenon as illustrating an "inference towards the antonym". The factors that determine the extent to which such an inference can take place are complex, and several proposals have been made (Horn 1989; Krifka 2007). Now, it has to be noted that given the context-dependency of the interpretation of gradable adjectives and their negation, we expect such inferences to be themselves context-dependent, so that the extent to which the inference can be made might itself be quite unclear. In a certain sense, *tall*, *not tall*, *short*, and *not short* all admit of a multiplicity of construals, which depend on a myriad of factors. When testing such inferences in an actual inferential task, we expect people's judgments to vary, and to be possibly gradient. For this reason, we will use the notion of ITA as reflecting a *gradient* phenomenon, and we will talk in terms of *strong* versus *weak* ITA effects.[3]

Rather than speaking of varying ITA strength, some authors use the expression "mitigation" (see, for instance, Giora et al. 2005; Fraenkel & Schul 2008). The idea is that using *not tall* rather than *short* mitigates the inference that the relevant person is short. Saying that the inference from *not tall* to *short* (inference towards the antonym) is strong then amounts to saying that the mitigation effect is weak. That is, a "weaker ITA effect" corresponds to a "stronger mitigation effect", and a "stronger ITA effect" corresponds to a "weaker mitigation effect".[4] In other words, ITA strength and mitigation correlate negatively. For the sake of clarity, we will consistently use the term ITA, as referring to a gradient phenomenon, and our focus is the *strength* of ITA for different adjectives.

It has been suggested in several works that ITA is stronger with positive adjectives than with negative adjectives. That is, *not tall* is more likely to be felt to imply *short* than *not short* is felt to imply *tall* (see, e.g., Krifka 2007). Krifka, partly building on Horn (1984, 1989), has offered an explicit theory of why this should be expected. One of the main goals of our study is to test this claim, as well as more fine-grained predictions that can be derived from Krifka's proposal, which concern the potential role of morphology in ITA strength differences between positive and negative adjectives.

In what follows we start by discussing the different notions of polarity (Section 1.2). We go on to discuss Krifka's (2007) theory of the negation of gradable adjectives, which

---

[3] Even if we adopt a somewhat idealized view where *not tall* is ambiguous between a weak, "logical" meaning as illustrated in Figure 1, and a stronger meaning equivalent to *short*, we would expect a certain amount of gradience: across adjectives and contexts, the probability of getting one or the other reading might vary. In an experimental setting, then, the average rate for the inference will vary across trials.

[4] See also Becker (2015), who provides corpus-based evidence for the mitigation effect of negation on adjectives, i.e., for the view that, in some cases, the negation of an adjective is not completely equivalent to the affirmation of its corresponding antonym.

forms the basis for our hypotheses (Section 1.3). First, following Krifka, positive adjectives should trigger a stronger ITA relative to negative adjectives. Second, with some auxiliary assumptions, Krifka's approach predicts greater ITA strength asymmetry for morphological antonymic pairs relative to non-morphological pairs. In Section 1.4, we review the available experimental evidence concerning these hypotheses.

In Section 2, we propose our own experimental investigation of the interpretation of negated antonymic adjectives. We introduce two new tests of polarity: the *not very* task and the *exclamation* task (Section 2.1). In a first experiment ITA strength is measured implicitly by having participants judge the pragmatic acceptability of sentences involving a negated adjective and its antonym. In a second experiment we collect explicit inferential judgments (Section 2.2). We find that both our hypotheses bear out: the adjectives classified as positives give rise to stronger ITA than the adjectives classified as negatives, and the asymmetry of ITA strength is greater for morphological pairs relative to non-morphological pairs (Sections 3 and 4).

1.2 Polarity

1.2.1 Notions of polarity

Members of antonymic adjectival pairs are generally thought of in terms of polarity (Cruse 1986: 246–257); polarity is a binary distinction concerning which member of a pair is the positive and which one is the negative member.

A number of quite distinct criteria have been offered to determine the polarity of an adjective. According to a morphological criterion, in a pair such as *correct/incorrect,* the positive member is the one which lacks the negative prefix (Horn 1989: 273–308). Thus the relevance of this criterion is limited to cases where the two members of a pair are transparently related from a morphological point of view. On top of this morphological criterion, at least three other distinct characterizations of polarity have been considered.

First, for *dimensional* adjectives, the positive member is the member whose associated scale is the one in which the relevant dimension is measured (von Stechow 1984; Bierwisch 1989: 79–82; Kennedy 1997: 61–62). For instance, *tall* is the positive adjective in *tall/short,* because its scale is the scale of heights: the taller one is, the higher the height, measured in some unit (on measurement scales, see Solt 2015). In other words, *tall* is positive because it refers to the possession of more centimetres (i.e., units of height) than *short* does.

Second, there is a notion of *evaluative* polarity associated with adjectives that inherently involve a subjective judgment of (un)desirability (Boucher & Osgood 1969; Horn 1989: 330–361; Paradis et al. 2012; Sassoon 2013). For such evaluative pairs, the positive member corresponds to the possession of the most desirable property. For example, *good* and *bad* are evaluative in virtue of their very meaning and *good* is the positive member because it is the more desirable of the pair. But even an adjective such as *tall,* which is not inherently evaluative, can acquire an evaluative use in a context where *being tall* is viewed as desirable.[5]

Finally, polarity is associated with the notion of *markedness*. In a pair of antonyms, the unmarked member (if there is one) is generally positive and the marked member

---

[5] Importantly, our use of the term *evaluative,* as referring to judgments of desirability, is to be distinguished from an entirely different but frequent use, where an adjectival construction is said to be evaluative if its truth-conditions depend on a context-dependent threshold. Brasoveanu & Rett (2016), for instance, define evaluativity as follows: "An adjectival construction is evaluative if and only if it conveys that the property associated with the adjective exceeds a relevant threshold". On the basis of this definition, the sentence *Peter is tall* is evaluative. According to our use of the term, however, the adjective good is semantically evaluative because its meaning relates to a judgment of desirability, while *tall* may or may not, depending on the context, convey such a judgment.

negative. Even though, as Lehrer (1985) proposes, markedness is best conceived of as a complex notion consisting of a set of several related characteristics, markedness is often associated with morphological negation. For instance, in a pair such as *happy/unhappy*, the marked member is the adjective with the negative prefix, i.e., *unhappy*. Other ways of measuring markedness have been proposed, however. One such test for markedness is discussed in Rett (2015): when used in an equative construction, a marked adjective triggers the presupposition that the adjective in its unmodified form applies to the items being compared. For the pair *tall/short*, for instance, there is a clear asymmetry between *tall* and *short* in terms of markedness, *short* being the marked member and *tall* the unmarked member. Whereas the sentence in (4) presupposes that both John and Peter are short in an absolute sense, (5) does not presuppose that John and Peter are tall in an absolute sense.

(4)   John is as short as Peter. → Both John and Peter are short.

(5)   John is as tall as Peter. ⇏ Both John and Peter are tall.

Thus, in the case of *tall/short*, all three notions of polarity result in the same conclusion regarding which member is positive and which one is negative. But this is not always the case, as we will see in the next section.

1.2.2 Discrepancies regarding polarity
The different notions of polarity are conceptually quite distinct and do not always map onto each other. Consider for instance the relation between evaluative polarity and markedness. Unlike pairs such as *tall/short*, in *happy/sad* both the evaluatively positive member *happy* and the evaluatively negative member *sad* appear to be marked, on the basis of Rett's (2015) equative test:

(6)   John is as happy as Peter. → Both John and Peter are happy.

(7)   John is as sad as Peter. → Both John and Peter are sad.

According to introspective judgments, the sentences in (6) and (7) both suggest that the predicates *happy* and *sad*, respectively, apply to both John and Peter. This example indicates that the unmarked/marked contrast does not always map onto the evaluative positive/negative contrast.[6]

  Neither is the positive member of an antonymic pair in the dimensional sense always unmarked. Take, for instance, the French pair *vieux/jeune* (*old/young*). As the contrast in (8)–(9) illustrates, the French pair *vieux/jeune* does not exactly correspond to the English pair *old/young*, in that *vieux* is as marked as *jeune*.

(8)   Jean est aussi jeune que   Pierre.
        Jean is   as   young as/than Pierre
        'John is as young as Peter.'
        → Both Jean and Pierre are young.

---

[6] This is also true when markedness is equated with morphological negation. A good example of a morphologically negative adjective that is evaluatively positive is the French *désintéressé* ('selfless')—even though the morphologically positive *intéressé* may be more limited in its uses.

(9)   Jean est aussi vieux que   Pierre.
        Jean is   as   old   as/than Pierre
        'John is as elderly as Peter.'
        → Both Jean and Pierre are old.

Yet, from a dimensional point of view, *vieux* (just as *old*) is the positive member of the pair.

   Finally, evaluative polarity does not necessarily correspond to polarity in the dimensional sense (see also Sassoon 2012 for similar remarks). While the members of the French pair *gentil/méchant* (*kind/mean)* are clearly positive and negative, respectively, in the evaluative sense, these adjectives are simply not dimensional. In French, both *gentil* and *méchant* have an associated noun (*gentillesse* and *méchanceté*) and therefore there is no single scale in which the relevant dimension is measured. Note that on the basis of the equative test both members of the pair are marked.

   With respect to evaluativity, it is worth noting that people do not always agree on what they consider positive or negative. For instance, while most people would agree that being *tall* is preferable to being *short*, jockeys are likely to disagree and consider *short* to be preferable. That is, the evaluative interpretation of gradable adjectives is context-dependent. There also are cultural influences on the interpretation of gradable adjectives (Reardon & Miller 2012). For instance, in some cultures being old might be more valuable than being young. Such issues about the variability in the interpretation of adjectives in terms of polarity suggest there is merit in using data-driven classifications of adjectives instead of a priori classifications that do not allow individual differences in polarity.

   Because the notion of adjective polarity is not completely clear from a conceptual point of view, we will operationalize it, rather than choose one notion and argue for its superiority over the others on theoretical grounds. In other words, we will classify members of adjectival pairs as positive or negative based on the performance of naïve subjects in two new linguistic tests. In this way we do not resolve the underlying conceptual issue, but we are at least using an objective and meaningful criterion to determine the polarity of adjectives. The reliance on subjects' performance instead of an a priori classification also allows for individual differences in polarity judgments.

1.3 Krifka's theory of the negation of gradable adjectives

Our hypotheses have their origins in Krifka's (2007) theory of the negation of gradable adjectives. Krifka (2007) mostly talks about morphological contrary adjectives, that is pairs one of whose members is derived from the other by adding a negative prefix, such as *unhappy* versus *happy*. However, he also explicitly mentions non-morphological pairs such as *few* versus *many* and *bad* versus *good*. He assumes Williamson's (1994) theory of vagueness, according to which sentences with vague predicates such as *tall* or *short* are always literally true or false. This is so even in so-called borderline cases, although the precise truth-conditions of such vague predicates might not be always known. Thus, in intermediary cases, one can be unable to tell whether someone is tall, even if one knows the person's precise height. For instance, if John is medium-sized, the truth-value of (10) is hard to determine.

(10)   John is not tall.

According to Krifka, *short* is synonymous with *not tall*, and *tall* is synonymous with *not short*. That is, the inference to the antonym (from *not tall* to *short* and from *not short* to *tall*) is viewed as an entailment. In other words, if the sentence *John is not tall* is true, then the sentence *John is short* is true too.

Krifka builds on Horn's work to account for the fact that, in practice, *not tall* is not necessarily perceived as equivalent to *short*. Horn's Q-Principle and R-Principle are conceived of as two antinomic principles that interact with and constrain one another (Horn 1989: 192–203). According to the Q-Principle, speakers should say as much as they can, given that they try to make a true contribution and that they respect the R-Principle according to which they should say no more than required. Krifka's line of explanation is the following. In agreement with Horn's (1984, 1989) R-Principle, the literal meaning of vague predicates such as *tall* and *short* will be enriched so that the uses of *tall* and *short* are restricted to stereotypical instances. In other terms, vague predicates tend to be used to describe *clear* cases, for which the uncertainty about the range of application of the predicate is immaterial. Furthermore, following Horn's Division of Pragmatic Labor, more complex expressions tend to be used to describe less stereotypical cases than equivalent expressions that are more simple (a principle that is viewed as a consequence of the Gricean maxim of manner). As a result, *not tall* will tend to be used to describe cases that fall under the literal meaning of *short* (since *short* and *not tall* have the same literal meanings), but which are less stereotypical instances of shortness. The end result is that *not tall* will tend to be applied to objects that are not as short as those that are typically described by the word *short*. Accordingly, *not tall* does not have exactly the same meaning as *short*, but means 'relatively short'.

Krifka's (2007) reasoning applies equally with *not short* (which is, literally, identical to *tall*), but with a twist. In line with what we will call the "Negative Adjectives Complexity Hypothesis" (NACH) endorsed by Büring (2007a, b), since *short* is defined formally as *tall* plus a negative prefix, *short* is inherently more complex than *tall*. According to NACH, negative adjectives are intrinsically more complex than their positive counterparts because they result from combining a negative morpheme with the corresponding positive adjective (see also Heim 2006, 2008). Thus, just as *unhappy* is (morphologically) more complex than *happy*, *short* is more complex than *tall* even though this difference of complexity is not morphologically transparent in this case. This means that the difference in complexity between *not short* and *tall* is greater than the one between *not tall* and *short*—in effect, *not short* can be viewed as *tall* prefixed by two negations (comparable to *not unhappy*, which Krifka discusses at length). Not only is *not short* more complex than *tall* because it contains one more word, the word *short* itself is more complex than the word *tall* (though *short* is less complex than *not tall*).[7] As a result of this, *not short* will be used for even less stereotypical cases than *not tall* is. In other words, the effect whereby the negation of an adjective tends to yield a weaker meaning than its antonym will be stronger with *short* than with *tall*. This intuition, that double negation gives rise to an attenuated sense of the corresponding antonym, can be traced back to Jespersen (1924: 332): "The two negative […] do not exactly cancel one another so that the result is identical with the simple [corresponding antonym]; the longer expression is always weaker" (quoted in Horn 1989: 296). Put in ITA terms, Krifka's (2007) analysis gives rise to the hypothesis that the inference to the antonym is expected to be stronger with *not tall* than with *not short*. Krifka thus predicts an effect of polarity on ITA strength: positive adjectives should give rise to a stronger ITA relative to negative adjectives. This is the first hypothesis we will test in this paper.

---

[7] Note that it is not entirely clear on this basis why *not tall* should be viewed as more complex than *short*, since *short* contains an abstract negative morpheme. Krifka assumes that a negative morpheme within a word creates less complexity than sentential negation, because the *number of words* is one of the factors that contributes to the complexity of an expression.

According to Horn (1989), another reason for violating the R-Principle is politeness considerations, such as the desire to avoid committing oneself to a face-threatening opinion (Brown & Levinson 1978). Instead of using a negative adjective, one prefers to negate its antonym (logic of euphemism). For instance, assuming that being short is generally undesirable, speakers might feel entitled to avoid *short* and use *not tall* instead. According to the logic of euphemism Horn assumes, it would be rational on the part of an interpreter to understand *not tall* as actually suggesting *short*. However, the interpreter need not understand *not short* as *tall*. That is, although Horn is not very explicit about it, because of politeness considerations he too seems to predict that positive adjectives in general should give rise to a stronger ITA than negative adjectives (for additional examples and discussion see Ducrot 1973; Israel 2004).

A further issue, which was not developed by Krifka (2007) but follows a similar rationale, concerns possible differences of ITA strength asymmetry between morphological antonymic pairs, in which members overtly differ in terms of morphological complexity in that the negative member contains an overt negative prefix (e.g., *unhappy* is more complex than *happy*), and non-morphological antonymic pairs, in which members possibly differ in terms of complexity, but if so not transparently (we will talk in this case of abstract complexity—*short* is more complex than *tall*).[8] Arguably, morphological complexity is more salient to interpreters than abstract complexity is, which, in turn, can influence how these expressions are processed. As a result, it might be expected that a negative adjective such as *short* is more easily reanalysed as being morphologically simple than a negative adjective such as *unhappy*. The connection between morphological antonyms (*happy/unhappy*) should be stronger than the connection between non-morphological antonyms (*tall/short*) because explicit negation is more salient to interpreters than implicit negation is (calling "implicit negation" a negation which is present in the underlying logical form but is not transparently expressed by a morpheme) (see Fillenbaum 1966). Likewise, the complexity difference between *happy* and *not unhappy* is more salient than that between *tall* and *not short* because two explicit negative morphemes are involved in the former case. Thus one might expect that the asymmetry with respect to ITA strength between the positive member and the negative member of the pair will be more pronounced with morphological pairs such as *happy/unhappy* than with non-morphological pairs such as *happy/sad* or *tall/short*.[9] This is the second hypothesis we will test in this paper.

1.4 Review of related work

To the best of our knowledge no study has addressed the effect of morphology on ITA strength. A handful of studies have looked at the effect of polarity on ITA strength, with conflicting results. Their comparison is hampered due to the use of different measures of ITA strength and the use of different notions of polarity. We nevertheless review these studies here.

Participants in Giora et al.'s (2005: 245–249) Experiment 3 were presented with pairs of sentences (*Sarit's dress was ugly/Sarit's dress was not pretty*) and asked to rate each sentence of the pair on a 7-point polarity scale (for example, *ugly – pretty*). The results showed that negated contrary adjectives and their antonyms were not rated similarly. The

---

[8] Our terminology here is purely descriptive and does not imply any strong view about the morphological make up of negative antonyms. It is possible that the best analysis of *short* is that it contains a covert negative morpheme. We conventionally say that a pair is *non-morphological* when the surface realization of the negative member does not involve an overt negative morpheme combined with the positive member.

[9] We do not need to assume here that when the negative prefix is overtly realized, the adjective is more complex than when the negative morpheme is not overtly realized. What we suggest, rather, is that a pragmatic reasoning that relies in part on the comparison of expressions in terms of complexity might be facilitated when the difference in complexity between expressions is made salient by overt morphology.

ratings for *ugly*, for instance, were dissimilar to those for *not pretty* (for a similar finding, see Paradis & Willners 2006). The discrepancy between negated contrary adjectives and their antonyms was the same regardless of the evaluative polarity of the adjectives. That is, the discrepancy between *ugly* and *not pretty* in the above examples was comparable to that between *pretty* and *not ugly* in the sentence pair *Sarit's dress was pretty/Sarit's dress was not ugly*. Thus, contrary to our prediction, Giora et al. found no effect of polarity on the strength of the ITA reading of negated adjectives.

An experiment by Colston (1999) provided partial support for the existence of ITA strength asymmetries between positive and negative adjectives. Colston asked participants to give their intuitions about what speakers mean by uttering English sentences such as *It is not hard* or *It is not soft* by placing an X on a scale with "very hard" and "very soft" as endpoints. Assuming an evaluative notion of polarity, Colston showed that the negation of a positive adjective was perceived as similar to the antonym of that adjective (*not soft* amounts to *hard*). By contrast, if the adjective was negative, its negation was not regarded as similar to the antonym (*not hard* does not amount to *soft*). The results of Colston (1999) thus support our predictions regarding the effect of polarity on ITA strength, but only when participants were first exposed to a short context inducing a positive expectation, such as: *Maria tried out the new sofa her husband bought them. When he asked her how it felt, she replied, "It isn't hard"*. When a negative expectation was induced, as in *Kim had to tell her father that she had wrecked his car. Her boyfriend asked her how that went and she said, "He wasn't happy."* no such asymmetry was found. Colston (1999) does not tell us whether a polarity effect would be found in minimal contexts that avoid inducing expectations.

Fraenkel & Schul (2008) investigated differences in ITA strength as a function of type (contrary versus contradictory) and markedness (marked versus unmarked). They collected judgments of meaning similarity between a sentence containing a negated adjective (11a) and a sentence containing the corresponding antonym (11b).

(11)   a.   John is not rich.
        b.   John is poor.

The results showed stronger ITA for contradictories such as *clear/vague* than for contraries such as *happy/sad*, and stronger ITA for unmarked adjectives such as *clear* and *good* relative to marked adjectives such as *vague* and *bad*. Inasmuch as markedness is an indication of polarity (see Section 1.2), these results are compatible with the prediction that negated gradable antonyms give rise to ITA strength asymmetries between positives and negatives.

In conclusion, using explicit inferential judgments, Colston (1999) demonstrated that, provided contextual information includes positive expectations, evaluative polarity can give rise to ITA strength asymmetry. Giora et al. (2005) too used explicit inferential judgments, but in decontextualized sentences did not find an effect of evaluative polarity on ITA strength. The latter finding is contradicted by Fraenkel & Schul's (2008) study, in which a stronger ITA effect was found for unmarked (positive) adjectives relative to marked (negative) adjectives using a meaning similarity rating task.

1.5 Intermediate conclusions

Let us take stock. According to Krifka (2007) and Horn (1989), it is reasonable to expect stronger ITA for positive versus negative adjectives. Following Krifka (2007) ITA strength asymmetry should also be greater for morphological antonymic pairs than for non-morphological pairs. To the best of our knowledge, no experimental study has been devoted to the latter hypothesis. Previous studies did not reach a uniform conclusion

regarding the former hypothesis and their comparison is hampered due to the use of different measures of ITA strength and the existence of different notions of polarity.

Our own contribution will be a) to evaluate the role of polarity, and the interaction of polarity and morphology, on ITA strength, b) to operationalize adjective polarity by means of linguistic tests, instead of resorting to a priori classifications based on a single interpretation of the polarity notion, and c) to introduce an indirect measure of ITA strength, which does not rely on speakers' metalinguistic intuitions about synonymy or meaning similarity. The next section is devoted to these new instruments.

2 Operationalizing polarity and inference

An important aim of this study is to evaluate Krifka's and Horn's predictions regarding the effect of polarity on the ITA phenomenon. In order to do so, we will not merely rely on an intuitive classification of adjectives into positive and negative adjectives (as in Colston 1999 and Giora et al. 2005), but we will exploit the fact that the polarity of an adjective correlates with the acceptability of certain types of sentences to determine adjective polarity. In that respect, our approach resembles Sassoon's (2012), who classified adjectives in positives versus negatives on the basis of participants' judgments. However, unlike Sassoon (2012), we will use an indirect measure based on the acceptability judgments of participants without ever mentioning "positive" or "negative" in the task itself. Likewise, because the acceptability of certain discourses and constructions is sensitive to inference patterns between expressions, we can use such discourses and constructions to probe implication relationships between various expressions. In the following section, we introduce two indirect measures of polarity and one indirect inference measure using some of the original French examples, used in the actual experiments.

2.1 Operationalizing polarity

In order to obtain an indirect polarity measure, we designed two linguistic tests: the *pas très* ('not very') test and the exclamation test.

2.1.1 Pas très

The first test builds on the intuition that evaluatively positive adjectives such as *gentil* ('kind') are more acceptable after *pas très* ('not very') than their antonyms, such as *méchant* ('mean') (see, for instance, Bolinger 1972: 116–118; Ducrot 1973; Horn 1989: 353–356, who hinted at this intuition).

(12) a. Jean n' est pas très gentil.
          Jean   is   NEG very kind
          'John is not very kind.'

      b. ??Jean n' est pas très méchant.
          Jean   is   NEG very mean
          'John is not very mean.'

This contrast is not specific to evaluative adjectives such as *gentil/méchant*, however. A similar contrast appears to hold for dimensional adjectives such as *grand/petit*, that are not inherently evaluative: (13a) is perfectly natural, whereas (13b) is not as natural in most contexts.

(13) a. Jean n' est pas très grand.
          Jean   is   NEG very tall
          'John is not very tall.'

    b. ??Jean n' est pas   très petit.
          Jean   is NEG very short
          'John is not very short.'

Now, note that (12b) can sound quite natural in a context where the question under discussion is whether it is the case that the predicate *very mean* applies to John. Note that all our experiments are context-free in the sense that all these sentences are presented in isolation. Thus what the discussion of these examples suggests is that negative adjectives need to be contextualized more than positive adjectives in order to be acceptable in the *not very* construction. This does not invalidate the claim that this linguistic test can be used to diagnose polarity.

2.1.2 Exclamations

Roche (2012) observed an interesting contrast in French exclamations involving *à quel point* ('the extent to which'). Namely, negated adjectives are quite acceptable in such exclamations if, from a dimensional or evaluative perspective, the adjective is positive, but less so if the adjective is negative, as illustrated in (14).

(14)   a.   C' est étonnant à   quel point Jean n' est pas   grand/gentil!
               it   is   surprising to what point Jean   is NEG tall/kind
               'It's surprising how John is not tall/kind!'

        b.   (?)C' est étonnant à   quel point Jean n' est pas   petit/méchant!
               it   is   surprising to what point Jean   is NEG short/mean
               'It's surprising how John is not short/mean!'

This contrast can serve as the basis for a second linguistic test (which we call the exclamation test) that will be used to confirm the robustness of the *pas très* test.[10]

2.2 Operationalizing inference: Aussi as probing implicatures

To test the extent to which *not-P* is interpreted as implying *Q*, where $<P, Q>$ are antonyms, we ran two distinct studies. In the first study, we choose to measure the strength of this inference by an *indirect test*, while the second study uses an explicit inferential task. We focus here on this indirect test.

   Our test takes advantage of the semantics and conditions of use of additive particles such as the English *too*, and the French *aussi*. Consider the following discourses:

(15)   a.   Mary lives in Paris. John too lives in France.
        b. ??Mary lives in France. John too lives in Paris.
        c. #Mary lives in France. John too lives in Asia.

While (15a) is perfectly natural, (15c) is completely deviant. (15b) is not natural, but might slightly improve in a context where it is known that in every country where Mary has lived or will live, she always chooses to live in the capital. There are many accounts of the semantics of *too* (see, e.g., Karttunen 1974; Heim 1990; van der Sandt & Geurts 2001; Kripke 2009), which differ regarding the details but share a common core. We do

---

[10] Roche's (2012) tentative explanation for this contrast is the following: *not tall,* in contrast with *not short*, can be reanalyzed as a degree predicate that can combine as a whole with a degree variable (just as *short*). In the case of *not short* (and in general *negation + negative scalar adjective*), the degree variable bound by the wh-operator is in the scope of negation, which creates a negative island violation. When the adjective is positive, the degree variable can occur *above* negation, so that there is no negative island violation.

not need to discuss them in any details here. We will assume the following: a sentence of the form *Y is Q too*, where *Y* is a referring expression and *Q* is a predicate, is felicitous only if previous discourse entails the existence of an individual *X* distinct from *Y*, such that *X is Q* is true. In a discourse of the form *X is P. Y is Q too*, the first sentence (*X is P*) is the one that can serve to satisfy the presupposition triggered by *too*, and this will be the case if and only if the material conditional *X is P → X is Q* holds. We can thus make the following assumption:

(16)   A discourse of the form *X is P. Y is Q too*, where *X* and *Y* are referring
         expressions, *presupposes* the material conditional *'X is P → X is Q'*.

Let us see how this captures the judgments given in (15). In (15a), the predicted presupposition is *Mary lives in Paris → Mary lives in France*. This presupposition is true by virtue of general common knowledge, so (15a) feels felicitous and does not trigger any remarkable inference. In (15b), the predicted presupposition is *Mary lives in France → Mary lives in Paris*, which is not true just by virtue of general common knowledge, hence the sense of deviance. But if we have further information about Mary, for instance if we know that in every country where she lives for a period of time, she lives in the capital of this country, the material conditional is true, and the sentence slightly improves (we think it would be more natural, in such a case, to simply say *Mary lives in Paris*). Finally, the predicted presupposition for (15c) is *Mary lives in France → Mary lives in Asia*, which, together with the knowledge that France is not in Asia and the information that Mary lives in France, is contradictory, hence the deviance of this example. From this generalization, we can derive another one:

(17)   The more *P* is felt to imply Q, the more a discourse of the form *X is P. Y is Q too*
         is felicitous.

The reason is the following. First, if *P* is ambiguous between a reading where it entails Q and one where it doesn't, the more the first reading is accessible, the more the conditional presupposition *X is P → X is Q* will be perceived as true. The extent to which *P* is felt to imply *Q*, in such a case, would track the extent to which the first reading is accessible. Second, even in the absence of ambiguity, we can reason as follows. When *P* does not logically entail *Q*, the presupposition *X is P → X is Q* needs to be *accommodated*, which we can assume comes at a cost, and the cost is higher when the presupposition that needs to be accommodated is implausible (Heim 1990; Singh et al. 2016). When *X is P* makes it *likely* that *X is Q*, the conditional presupposition *X is P → X is Q* is itself more likely to be true than when *X is P* does not make it likely that *X is Q*.

The generalization in (17) can be illustrated by the following pair:

(18)   a.   Mary writes books. Peter too is an intellectual.
         b. #Mary writes books. Peter too likes movies.

While *writing books* does not logically entail *being an intellectual*, the inference from *writing books* to *being an intellectual* is easier to accept, given common knowledge, than the inference from *writing books* to *liking movies*. Given this generalization, the extent to which *not P* is felt to imply *Q*, where *P* and *Q* are antonyms, can be tested by measuring the acceptability of a discourse of the form *X is not P. Y is Q too*. We illustrate in (19) the kind of discourses that we will test in French to compare the ITA strength of antonyms such as *grand* and *petit*.

(19) a. Paul n' est pas grand. Pierre aussi est petit.
          Paul   is   NEG tall   Pierre too   is   short
          'Paul is not tall. Peter is short too.'

      b. Paul n' est pas petit. Pierre aussi est grand.
          Paul   is   NEG short Pierre too   is   tall
          'Paul is not short. Peter is tall too.'

If positive adjectives give rise to stronger ITA than negative adjectives do, sequences such as (19a) will be more acceptable than discourses such as (19b). The two linguistic tests of polarity and the indirect measure of inference are put to use in Experiment 1 to test these hypotheses.

3 Experiment 1
3.1 Method
3.1.1 Participants
We recruited 37 native speakers of French (29 females; mean age 21.9 years, standard deviation 2.4, range 17–27) during classes at the Université libre de Bruxelles. None of them had a degree in linguistics or had previous experience with the experimental design.

3.1.2 Materials
All of the adjectives in our study were context-dependent, relative adjectives, according to Kennedy (2007) and Kennedy & McNally's (2005) definition: the standard according to which one determines that "X is ADJ" is determined by the context of utterance (as opposed to the scale for absolute adjectives). We tried to include as many canonical antonymic pairs as possible (cf. Jones et al. 2012), resulting in 13 morphological and 14 non-morphological pairs of French adjectives. Five pairs consisted of contradictory adjectives (*even/odd, true/false, certain/uncertain, accurate/inaccurate, possible/impossible*). With the exception of *even/odd*, the adjectives of these five pairs could be coerced into contraries (cf. Section 1.1). Table 1 lists the adjectival pairs, along with their English translation and a classification in terms of morphology.

In Table 1 the left adjective of a pair is always the positive member, while the right adjective is the negative member, according to the following a priori classification. In a morphological pair, adding to the (morphologically) unmarked member a negative prefix forms the (morphologically) marked member; the marked member can be decomposed into the prefix and the adjective.[11] We considered the unmarked member as the positive member of a pair, and the marked member as the negative member. For non-morphological pairs, we determined the positive and negative members of the pairs by relying on the notion of dimensional polarity. For instance, we considered *old* as the positive member of the pair *old/young* because it ranks higher on the scale of age than *young* does. For the pairs that are inherently evaluative, we used the notion of evaluative polarity: accordingly, *kind* is positive, and *mean* is negative. We thus resorted to three different notions to determine a priori the polarity of the members of the antonymic pairs. This is not unexpected given the fact that the notion of polarity is a complex one (see Cruse 1986: 246–257 and Section 1.2 for a discussion).

---

[11] Strictly speaking, this is not the case for *facile/difficile*. However, both adjectives are phonologically similar and their historical morphological relation may still be available to speakers. We therefore decided to classify that pair as morphological, in contrast to the non-morphological pair *facile/compliqué*.

Table 1: A priori classification of the adjectives.

adjectival pair	English translation	morphology
certain/incertain	certain/uncertain	morphological
chanceux/malchanceux	lucky/unlucky	morphological
exact/inexact	accurate/inaccurate	morphological
facile/difficile	easy/difficult	morphological
heureux/malheureux	happy/unhappy	morphological
intéressant/inintéressant	interesting/uninteresting	morphological
juste/injuste	fair/unfair	morphological
pair/impair	even/odd	morphological
poli/impoli	polite/impolite	morphological
possible/impossible	possible/impossible	morphological
satisfaisant/insatisfaisant	satisfactory/unsatisfactory	morphological
sympathique/antipathique	friendly/unfriendly	morphological
utile/inutile	useful/useless	morphological
bon/mauvais	good/bad	non-morphological
calme/agité	quiet/restless	non-morphological
chaud/froid	hot/cold	non-morphological
facile/compliqué	easy/complicated	non-morphological
fort/faible	strong/weak	non-morphological
gentil/méchant	kind/mean	non-morphological
grand/petit	tall/short	non-morphological
joyeux/triste	happy/sad	non-morphological
long/court	long/short	non-morphological
poli/grossier	polite/rude	non-morphological
riche/pauvre	rich/poor	non-morphological
satisfaisant/frustrant	satisfactory/frustrating	non-morphological
vieux/jeune	old/young	non-morphological
vrai/faux	true/false	non-morphological

3.1.3 Procedure

The experiment comprised three blocks presented in a sequential order, corresponding to the indirect inferential task with *aussi*, the *pas très* test, and the exclamation test (within-participants design).[12] We included a pause between each block. None of the blocks included training items. Experimental items and filler items (see Section 3.1.3.1) were randomly displayed within blocks, and the randomized order was different for every participant.

The instructions were identical for the three blocks. The participants were asked to indicate on a 7-point rating scale the extent to which they judged one or a sequence of two sentences to be relatively natural ("*plutôt naturelle*") or, on the contrary, relatively odd ("*plutôt bizarre*"). For each item participants had to select as a response one option, ranging from "completely odd" to "odd", "relatively odd" on the left, to "neutral" as intermediate position, and "relatively natural", "natural", and "completely natural" on the right. After each response, the next item was automatically displayed. The participants were told that they should respond as spontaneously as possible, and that making a judgment should not take them more than a few seconds. The word-by-word instructions are available on the Open Science Framework, along with the raw data and the results of the analyses.[13]

---

[12] The experiment actually included a fourth block in which we attempted to collect indirect measures of whether an adjective is inherently evaluative or not. However, the linguistic test we used was not satisfactory.

[13] https://osf.io/skp4t/?view_only=ed707a01aaaf4c09a8b6def00276d6f9.

3.1.3.1 Inference block

In the inference block, the participants were asked to evaluate whether a sequence of two sentences such as (19) was natural or not. A total of 54 such sequences were created, involving the two members of the 27 antonymic pairs either in the order shown in (19a) or in the reversed order (19b).

(19a)   (repeated)
       Paul n' est pas grand. Pierre aussi est petit.
       Paul   is NEG tall   Pierre too   is   short
       'Paul is not tall. Peter is short too.'

We wanted to ensure that the participants would actually read the sentences and that they would make their pragmatic acceptability judgments on the basis of a relation between the two sentences of a sequence. We thus wanted to prevent the participants from responding on the basis of only one sentence instead of the whole sequence.

Three types of fillers were therefore used (#128). The first type (#54) consisted in obviously acceptable sequences of sentences. An example is the sequence in (20), where the inference from *grande* (tall) in the first sentence to *grande* (tall) in the second one is trivially valid.[14]

(20)   Juliette est grande. Marie aussi est grande.
       Juliette is tall.FEM Marie too   is   tall.FEM
       'Juliette is tall. Mary is tall too.'

In sequences such as (20), the second sentence triggers the presupposition that someone different from Marie is *tall*; the antecedent of this proposition is asserted with the utterance of the first sentence (Winterstein 2011; see also Section 2.2).

We also used a second type of fillers (#54) consisting in trivially invalid inferences, as (21) illustrates: the negation of *grand* (tall) cannot be understood as the affirmation of *grand* (tall) for the reason that the use of *aussi* (too) cannot be accommodated.

(21)   Pierre n' est pas grand. Thierry aussi est grand.
       Pierre   is NEG tall   Thierry too   is   tall
       'Peter is not tall. Thierry is tall too.'

We also considered these two types of fillers as baselines relative to which the participants would make their judgments for sequences with a more debatable acceptability.

We used a third type of fillers (#20) involving different adjectives, and which consisted in inferences with debatable validity, such as (22) and (23).

(22)   Cette maison est immense. Celle-ci   aussi est grande.
       that house is huge   This one too   is   big.FEM
       'That house is huge. This one is big too.'

(23)   Cette maison est grande. Celle-ci   aussi est immense.
       that house is big.FEM This one too   is   huge
       'That house is big. This one is huge too.'

---

[14] However, it should be pointed out that, from a stylistic point of view, the repetition of the adjective is not very natural in French—this could affect the participants' judgments.

The direction of entailment between *grande* ('big') and *immense* ('huge') is from *immense* to *grande*. What is *huge* is necessarily *big*, but what is *big* may not necessarily be *huge*. This is why the inference in (22) can be considered valid, and this sequence is rather natural. By contrast, it is much less obvious that the inference involved in (23) is valid: *grande* does not entail *immense*, which is a reason why the sequence would be less acceptable than that in (22).

3.1.3.2 *Pas très* block

In the *pas très* block, the participants were asked to evaluate whether a sentence such as (13a) was natural or not (see Section 2.1.1).

(13a)　　(repeated)
　　　　Jean　n'　est　pas　très　grand.
　　　　Jean　　is　NEG　very　tall
　　　　'John is not very tall.'

A total of 54 sentences were presented, one for each of the 2 × 27 adjectives from Table 1. No fillers were used.

3.1.3.3 Exclamation block

In the exclamation block, the participants were asked to evaluate whether a sentence such as (14a) was natural or not (see Section 2.1.2).

(14a)　　(repeated)
　　　　C'　est　étonnant　à　quel　point　Jean　n'　est　pas　grand/gentil!
　　　　it　is　surprising　to　what　point　Jean　　is　NEG　tall/kind
　　　　'It's surprising how John is not tall/kind!'

A total of 54 sentences were presented, one for each of the 2 × 27 adjectives from Table 1. No fillers were used.

3.2 Predictions

The acceptability judgments in the *pas très*, the exclamation, and the inference blocks were made on 7-point scales ranging from "completely odd" to "completely natural". These judgments were recorded in scores from 1 to 7, the higher the score, the higher the pragmatic acceptability of the sentence or pair of sentences.

　　The variable of polarity was determined on the basis of the relative acceptability scores for the members of a pair in the *pas très* and exclamation tasks, with the higher scoring member being the positive member and the lower scoring member the negative one (see Section 3.3.1 below for details).

　　ITA strength was quantified by the scores for the experimental items in the inference task, for which higher acceptability scores indicate higher ITA, such as in (19a).

(19a)　　(repeated)
　　　　Paul　n'　est　pas　grand.　Pierre　aussi　est　petit.
　　　　Paul　　is　NEG　tall　　Pierre　too　　is　　short
　　　　'Paul is not tall. Peter is short too.'

We predicted that the participants' acceptability judgments would reflect a stronger ITA effect for positive versus negative adjectives, with a higher ITA strength asymmetry for morphological versus non-morphological pairs. As discussed in Section 1.3, we expect,

on the one hand, a stronger ITA effect for negated positive adjectives than for negated negative adjectives. On the other hand, the complexity difference between the positive member and the negative member should be more salient to speakers/interpreters in morphological antonymic pairs such as *happy/unhappy* than in non-morphological pairs such as *tall/short*, so that this contrast in ITA strength should be stronger with morphological pairs.

3.3 Results

3.3.1 Polarity results

We tested the reliability of the *pas très* judgments and found a Cronbach's alpha of .98. For the exclamation judgments Cronbach's alpha was .95. These values signal very high agreement among the participants within a task and thus provide an indication that we can use the average judgments on each of the tasks to determine whether an adjective is positive or negative. The average *pas très* and exclamation judgments also correlated highly (r = .86). On the basis of these results, we decided to make a dichotomized classification of the members of the adjectival pairs in positives/negatives. That is, we wanted to determine the polarity of the members of antonymic pairs on the basis of their average acceptability score—the member of the pair with the highest score being the *positive* member (Section 2.1). We first discarded four pairs, for which the direction of the difference was not identical in the two tests (the last four pairs of Table 2, i.e., *quiet/restless*, *easy/complicated*, *easy/difficult*, and *old/young*).

To determine whether the acceptability judgments for the two members of each of the remaining pairs differed significantly, we carried out 23 Wilcoxon signed rank tests, which posed a problem of multiple comparisons. To address this problem, we used the Holm-Bonferroni correction method and adjusted the alpha value of the signed rank test from .05 to .05/(k-i + 1), where k is the number of comparisons (pairs) and i indexes the pair according to the extremeness of the corresponding p-value (i = 1 is most extreme). According to the results of these tests, 2 out of the 23 remaining pairs were discarded because the difference between the two members of the pair was not significant in both tests (*even/odd* and *hot/cold*). The classification of the members of the remaining 21 pairs proved to be identical to the a priori classification that can be found in Table 1. This finding adds to the validity of the proposed tests of polarity. The analyses for the inference task with *aussi* will only concern the 21 pairs for which the two polarity tests provide a significant difference between the scores for the two members of the pairs. Note, however, that similar results were obtained in comparable analyses in which all pairs were employed and polarity was determined based on the relative mean scores of the *pas très* or the exclamation test, and in "participant-specific" versions of the analyses in which we determined positivity versus negativity on the basis of the individual participants' judgments instead of the average.

3.3.2 Inference results

Before focusing on the results concerning the effect of polarity and morphology on ITA strength, we first present the results for the different sorts of fillers in Figure 2.

Fillers such as *John is tall. Peter is tall too*, which we considered being completely valid, gave rise to the highest acceptability scores (mean of 6.44 on the 7-point scale). By contrast, the invalid fillers consisting in contradictions, such as *John is not tall. Peter is tall too*, resulted in the lowest scores (2.16). Between these two extremes, the two sorts of debatable fillers—"valid" ones, such as *That house is huge. This one is big too* and and"invalid" ones, such as *That house is big. This one is huge too*—gave rise to intermediate scores, with the acceptability scores for the former (5.07) being higher than those

Table 2: Average results on the two polarity tests for positive and negative members of antonymic pairs (means, standard deviations, and results of the Wilcoxon signed rank tests).

| pair	pas très positive mean (SD)	pas très negative mean (SD)	Wilcoxon signed rank test V	Wilcoxon signed rank test p-value	pair	exclamations positive mean (SD)	exclamations negative mean (SD)	Wilcoxon signed rank test V	Wilcoxon signed rank test p-value
fair/unfair	6.51(0.96)	2.62(1.60)	595	<.001	fair/unfair	5.32(1.51)	2.57(1.21)	463	<.001
polite/impolite	6.62(0.49)	2.73(1.77)	593.5	<.001	polite/impolite	5.62(1.48)	3.19(1.54)	484	<.001
satisfactory/unsatisfactory	6.27(0.99)	2.73(1.76)	560	<.001	satisfactory/unsatisfactory	5.62(1.64)	3.08(1.53)	533.5	<.001
useful/useless	6.3(1.08)	2.89(1.75)	588	<.001	useful/useless	5.11(1.47)	3.22(1.42)	396	<.001
friendly/unfriendly	6.57(0.83)	2.94(1.87)	528	<.001	friendly/unfriendly	5.73(1.19)	3.65(1.73)	445	<.001
satisfactory/frustrating	5.92(1.42)	2.73(1.66)	528	<.001	satisfactory/frustrating	4.46(1.82)	2.49(1.26)	496	<.001
polite/rude	6.54(0.65)	3.62(1.88)	493	<.001	polite/rude	5.05(1.68)	3.27(1.22)	417	<.001
happy/unhappy	6.24(1.06)	3.27(1.71)	597.5	<.001	happy/unhappy	5.19(1.51)	3.11(1.61)	491	<.001
strong/weak	6.49(0.99)	3.70(1.79)	435	<.001	strong/weak	4.97(1.8)	3.22(1.56)	386.5	<.001
good/bad	6.62(0.76)	3.97(2.14)	378	<.001	good/bad	5.38(1.74)	3.76(1.59)	331	<.001
interesting/uninteresting	6.08(1.8)	2.65(1.69)	507	<.001	interesting/uninteresting	5.13(1.68)	3.32(1.41)	395	<.001
tall/short	6.73(0.45)	4.67(1.94)	325	<.001	tall/short	4.94(1.6)	3.32(1.56)	324.5	<.001
kind/mean	6.51(0.65)	4.76(1.80)	396	<.001	kind/mean	4.81(1.63)	3.43(1.64)	343	<.001
rich/poor	6.08(1.3)	4.38(1.95)	300	<.001	rich/poor	4.51(1.68)	3.13(1.68)	364	<.001
long/short	6.30(1.15)	4.67(1.76)	291	<.001	long/short	5.43(1.76)	3.84(1.82)	363	<.001
certain/uncertain	4.35(1.95)	2.49(1.63)	428	<.001	certain/uncertain	5.19(1.73)	3.92(1.77)	293	<.001
lucky/unlucky	6.27(1.35)	2.94(1.53)	624	<.001	lucky/unlucky	5.27(1.85)	3.51(1.99)	250.5	<.001
accurate/inaccurate	5.27(1.89)	2.03(1.12)	630	<.001	accurate/inaccurate	4.38(1.72)	3.05(1.60)	308.5	<.001
true/false	4.19(2.07)	2.97(1.69)	267.5	<.001	true/false	4.76(1.85)	3.40(1.99)	348	<.001
happy/sad	6.13(1.03)	4.97(1.66)	260.5	<.001	happy/sad	4.73(1.59)	3.73(1.77)	156.5	<.001
possible/impossible	2.86(1.49)	2.05(1.31)	200	<.01	possible/impossible	3.81(1.61)	3.27(1.59)	176.5	<.05
cold/hot	5.49(1.72)	4.92(1.80)	182.5	<.05	cold/hot	4.38(1.8)	3.73(1.73)	161	>.05
even/odd	1.86(1.46)	1.76(1.40)	44	>.05	even/odd	1.76(1.4)	1.51(1.09)	13.5	>.05
quiet/restless	6.08(1.32)	5.62(1.67)	136	<.05	quiet/restless	4.08(1.83)	4.72(1.68)	53.5	<.05
easy/complicated	6.32(1.18)	6.61(0.6)	25	>.05	easy/complicated	6.08(1.34)	5.65(1.20)	130	<.05
easy/difficult	5.54(1.74)	6.38(1.16)	24	<.01	easy/difficult	5.78(1.27)	5.51(1.59)	112	>.05
old/young	6.11(1.35)	5.89(1.49)	76.5	>.05	old/young	4.19(1.76)	4.22(1.66)	102	>.05
average	5.79(1.23)	3.74(1.63)			**average**	4.88(1.63)	3.51(1.57)		
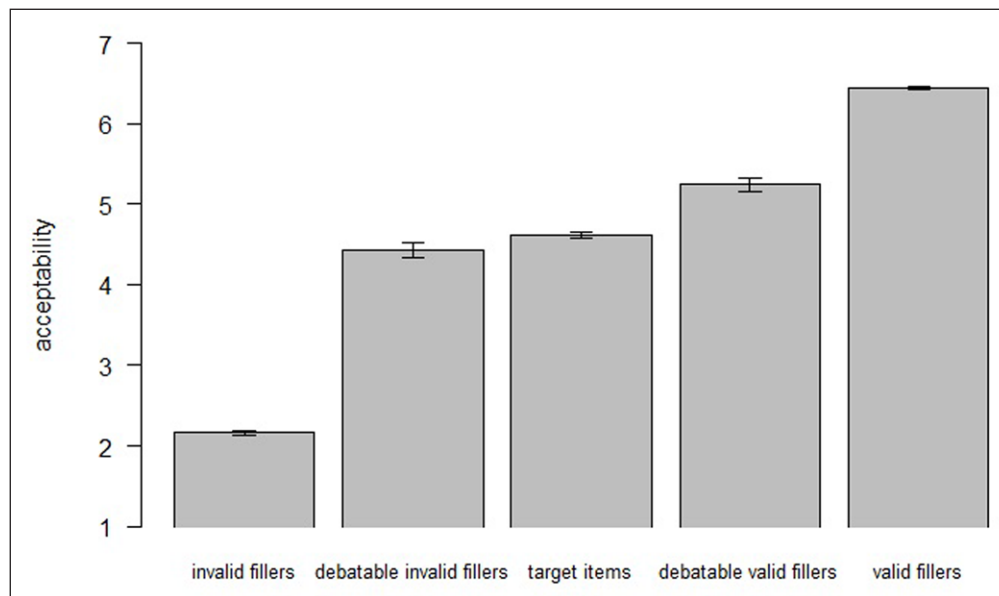
Figure 2: Average scores for the four types of fillers and the target items in the inference task.

of the latter (4.35). Taken together, these results suggest that, as we expected, the participants judged the acceptability of the combination of two sentences, rather than the acceptability of only one sentence. The ratings for the target items (4.61)—averaged across positive/negative and morphological/non-morphological—were higher than for the debatable invalid entailments, but lower than for the debatable valid ones. Taken together, the results for the fillers and the target items support the view that the ITA is a gradient effect.

To determine whether the effects of polarity and the interaction between polarity and morphology on inference were statistically significant, the acceptability judgments in the inferential task were fitted using a linear mixed effects model. The analyses were carried out in R (version 3.2.2 for Windows) (R development core team, 2015), employing the lme4 package (Bates & Sarkar 2007). We report the t-statistic and treat it as a z-statistic to derive p-values following Barr et al. (2013).

The inference scores were regressed on two predictors: one predictor called Polarity, which is a binary variable indicating whether an adjective is classified as the positive or negative member of an antonymic pair, and one predictor called Morphology, indicating whether the adjective is part of a morphological or non-morphological antonymic pair. The random part of the model included random participant and pair intercepts, by-pair and by-participant random slopes for Polarity, by-participant random slopes for Morphology, and by-participant random slopes for Polarity and Morphology. The fixed part of the model contained main effects of Polarity and Morphology and a Polarity x Morphology interaction. The results showed a main effect of Polarity on inference scores ($t = 8.94$; $p < .001$). Thus, in our experiment, adjectives classified as positives give rise to higher inference scores than adjectives classified as negatives do (see Appendix 1 for an overview of the mean scores and standard deviations for the individual adjectives). The effect of Morphology was not significant ($t = 1.45$; $p = .15$). However, we found the interaction between Morphology and Polarity to be statistically significant ($t = -4.91$; $p < .001$), indicating that the inference score asymmetry between positives and negatives was significantly stronger for morphological pairs relative to non-morphological pairs. This interaction is illustrated in Figure 3.
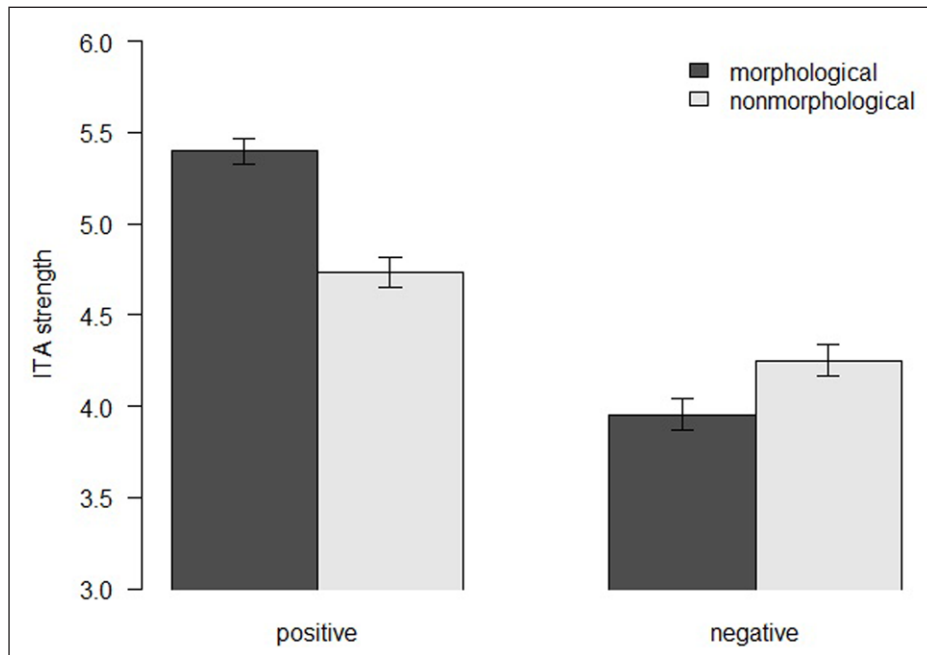
Figure 3: Mean inference scores according to morphology and polarity (error bars represent standard deviations from the means) in Experiment 1.

3.4 Discussion

Our two hypotheses are confirmed by the inference task results. First, the ITA was stronger for positive adjectives than for negative adjectives. Second, ITA strength asymmetry between positives and negatives was stronger for morphological pairs than for non-morphological pairs, as shown by the significant interaction between Morphology and Polarity: when an antonymic pair is morphological, the difference in ITA strength between the negation of the negative member (e.g., *not unhappy*) and that of the positive member (e.g., *not happy*) is larger. These results suggest that overt (i.e., morphological) complexity plays a greater role than covert (i.e., "abstract") complexity (e.g., *short* is more complex than *tall* because it includes a covert negative prefix) in ITA effects. As we suggested in Section 1.3, a possible reason for this is that morphological complexity contributes more to the comprehension processes than abstract complexity because the former is more salient to interpreters. Double negation and, in particular, double morphological negation, results in a weakening of ITA effects.

At this stage, it is worth considering a possible alternative explanation for our results. Recall that, in the inference task, the premise of the target items was always a negative sentence such as *Peter is not short*. We just saw that in this task negative adjectives received lower inference scores than positive adjectives did. Our linguistic tests of polarity (with *pas très* and exclamations), in which positive adjectives gave rise to higher acceptability scores than negative adjectives, both included an explicit marker of negation. One might thus argue that our results in the inference task reflect the fact that, in general, negative adjectives are dispreferred when they are under the scope of a linguistic negation. According to such a view, the difference between positive and negative adjectives in the inference task would not necessarily reflect differences in ITA strength.

Thus, according to this account, our result according to which positive adjectives are rated higher than negative adjectives in the inference task could be due to the fact that, in general, negative adjectives are less acceptable than their positive counterparts in linguistics contexts containing a negation. Sassoon (2012) provides empirical evidence for such an asymmetry of acceptability.

In order to evaluate the plausibility of this alternative account, we conducted an additional experiment in which participants were asked to make inferences towards a scale (in line with Fraenkel & Schul 2008; see Section 1.4). If the same effects were to hold for the data obtained with this new inference task that does not rely on acceptability judgments, the alternative account does not hold.

4 Experiment 2

4.1 Method

4.1.1 Participants

The participants in this online experiment were 25 volunteers recruited among students and graduate students at the Université libre de Bruxelles (18 females, mean age 23.5, standard deviation 4.3, range 20–38). All of them were native speakers of French. None of them had previous experience with this experimental design or any expertise with the issue under investigation.

4.1.2 Materials

The experimental items consisted of 51 negative statements such as (24):

(24)   Peter is not tall.

In addition to the 51 negative statements, we included 51 fillers consisting of the affirmative counterparts of the "inference items", such as (25):

(25)   Peter is tall.

The number 51 corresponds to the number of different adjectives used in Experiment 1. The adjectives *poli*, *facile*, and *satisfaisant*, which have two antonyms (a morphological one, i.e., *impoli*, *difficile*, and *insatisfaisant*, and a non-morphological one, i.e., *grossier*, *compliqué*, and *frustrant*) were used only once in this experiment. The classification of these items into morphological/non-morphological and positive/negative was identical to that used for Experiment 1 (see Sections 3.1.2 and 3.3.1).

4.1.3 Procedure

There was no training block provided. The experimental and filler items were presented individually in a single block. The presentation of the items was randomized for every participant.

Below each sentence such as (24)–(25), a graded scale was positioned, ranging from the negated/affirmed adjective on the left (*tall*) to the antonym of the negated/affirmed adjective on the right (*short*). The participants were asked to judge the subject of the sentence by positioning the cursor on the scale and clicking to validate their response. We stressed that they should respond as spontaneously as possible and that there were no right or wrong responses. After clicking on the judgment bar, the participants automatically moved on to the next statement. The word-by-word instructions are available on the Open Science Framework, along with the raw data and the results of the analyses.[15]

4.2 Predictions

We expected the participants to select the right-side of the scale for experimental items. The more to the right the participants position the cursor in response to a negated adjective, the more ITA they display. ITA strength was quantified by awarding clicks on the

---

[15] https://osf.io/skp4t/?view_only=ed707a01aaaf4c09a8b6def00276d6f9.

scale a score in the interval [0, 1] proportional to the distance from the left end of the scale. Clicks on the far left end of the scale were thus awarded a score of zero, while clicks on the far right end of the scale were awarded a score of 1. The higher the score, the higher the ITA. As in Experiment 1, we predicted that the participants' judgments on the scale would reflect a stronger ITA effect for positive versus negative adjectives, with a stronger ITA strength asymmetry for morphological versus non-morphological pairs.

4.3 Results

The mean scores obtained in Experiment 2 are provided in Figure 4. Appendix 1 lists the mean scores and standard deviations for the individual adjectives. Both in the Figure and in the Appendix, higher scores reflect stronger inferences towards the antonym.

To determine the statistical significance of the differences between these means, the ratings from the new inference task were fitted using the same linear mixed effects model we used to analyse the data from the original inference task. The results indicate a main effect of Polarity on explicit inference ratings ($t = 6.59$; $p < .001$). Thus, in both our experiments, adjectives classified as positives give rise to stronger ITA than adjectives classified as negatives do. As in Experiment 1, the results do not show a main effect of Morphology on inference ($t = 0.89$; $p = .37$). The results do show that the ITA asymmetry between positives and negatives is significantly stronger for morphological pairs relative to non-morphological pairs, just as it was in Experiment 1: we found a significant interaction of Morphology and Polarity on inference ($t = -3.18$; $p < .01$).

As was the case for Experiment 1, the main effect of Polarity and the interaction of Polarity and Morphology on inference scores were also significant in analyses taking into consideration all pairs and in which polarity was determined based on the relative mean scores of the *pas très* or the exclamation test, as well as on the basis of the individual participants' judgments instead of the average (see the Open Science Framework for these results).[16]
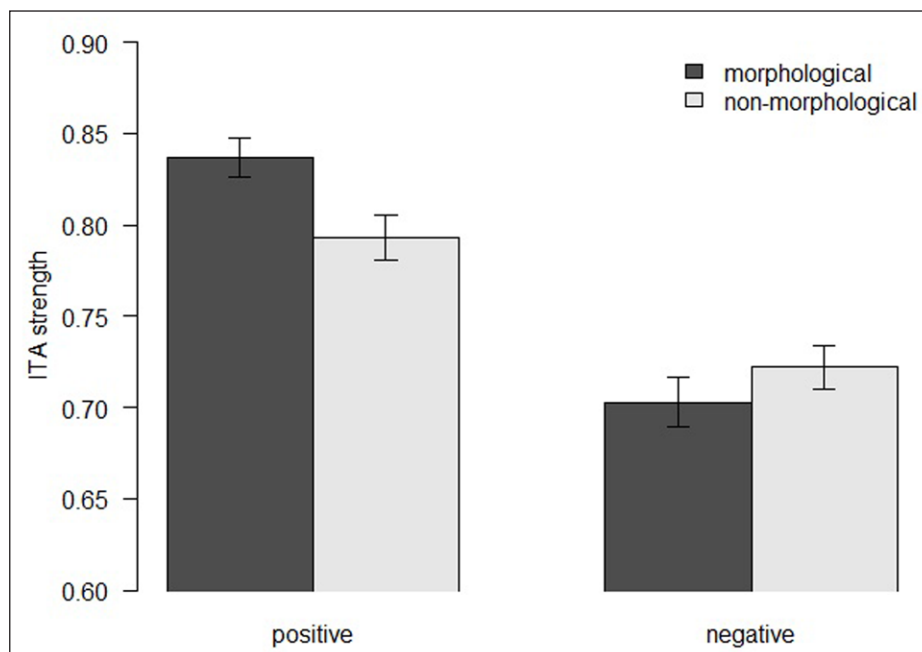

Figure 4: Mean inference scores according to morphology and polarity (error bars represent standard deviations from the means) in Experiment 2.

---

[16] https://osf.io/skp4t/?view_only=ed707a01aaaf4c09a8b6def00276d6f9.

4.4 *Discussion*

We considered an alternative account of our results, according to which the difference between the inference scores for positive versus negative adjectives in our indirect acceptability task (*aussi* task) does not necessarily reflect differences in ITA strength, but rather decreased pragmatic acceptability of negated negative adjectives. The results of Experiment 2 show that the alternative account does not apply. We obtained similar effects in an inference task where pragmatic acceptability of negation was not in play.

5 General discussion

Our experimental studies on ITA arising from negated gradable adjectives yield several interesting results.

In Experiment 1 we found stronger ITA for the adjectives classified as positives relative to the adjectives classified as negatives using an indirect inference task consisting in linguistic acceptability judgments (*aussi* task). This result confirmed Krifka's (2007) and Horn's (1989) predictions about the role of polarity on ITA strength. We found this effect with polarity determined by means of two new tests (the *pas très* test and the exclamation test), which measure polarity in a linguistically meaningful way and allow one to determine, in a bottom-up, data-driven manner, the polarity of the members of antonymic pairs. The polarity classification resulting from these tests was identical to an independent classification of polarity based on established grounds. In other words, the effect of polarity on ITA strength holds whether or not one considers the *pas très* and the exclamation test appropriate measures of polarity.

We interpreted the results of the indirect inference task in terms of stronger ITA for the positive adjectives relative to the negative adjectives. Another possible explanation of these results was discussed according to which the reason why positive adjectives give rise to higher inference scores in Experiment 1 is that, in general, negative adjectives are less felicitous than positive adjectives in linguistic environments with a negation (Sassoon 2012). We rejected this explanation on the basis of data from Experiment 2 consisting of explicit inference ratings instead of acceptability ratings. The results of the explicit inference task confirmed the results obtained with the indirect inferential task. This strongly suggests that adjective polarity is responsible for the differences in ITA. This finding contradicts that of Giora et al. (2005), but is in line with Colston (1999) and Fraenkel & Schul (2008).

In line with our developments of Krifka's (2007) theory, we also established both in Experiment 1 and in Experiment 2 an interaction between polarity and morphology on ITA effects: ITA strength asymmetry was more pronounced for morphological pairs than for non-morphological pairs.

A plausible explanation of these findings can be found in Horn's Division of Pragmatic Labor, following Krifka (2007). According to this explanation, negated negative adjectives, such as *not impolite*, or *not short*, are doubly negative. Therefore, relative to their affirmed positive counterparts, such as *polite* and *tall*, which constitute the simplest expression available, they are perceived as more complex by interpreters and their use is restricted to non-stereotypical cases. This explains the fact that ITA is stronger for positive adjectives than for negative adjectives. Furthermore, the driving factor of this contrast, namely the difference in complexity between positive and negative adjectives, is more pronounced and more salient when the antonymic pair is morphological. Thus the contrast between positive and negative adjectives is stronger in the case of morphological pairs.

6 Conclusion

In this paper, we investigated the various interpretations of negated adjectives. A sentence involving a negated adjective (e.g., *Peter is not tall*) can be understood as meaning either that Peter is not tall tout court, or that Peter is rather short ("inference towards the antonym"—ITA). In an experiment with French material, we tested predictions of Krifka's (2007) theory. First, we predicted that the ITA should be stronger for positive adjectives than for negative adjectives. Second, we predicted that the ITA strength asymmetry should be stronger for morphological pairs relative to non-morphological pairs.

To test these predictions, we determined the polarity of the members of antonymic pairs according to their average acceptability score in two linguistic tests. Interestingly, the objective classifications of adjectives in positives versus negatives we obtained on the basis of the scores in these tests confirmed the initial classification based on commonly accepted criteria: the pairs for which the difference between positives and negatives was significant in our two tests matched the a priori classification. The results of these two linguistic tests thus strongly suggest that they can be used to determine polarity. The data of two inference tasks confirmed that the adjectives classified as positives give rise to stronger ITA than their counterparts classified as negatives. Both inference tasks also provided support for our predictions about the effect of morphology on ITA strength asymmetry.

In this paper we started with a discussion of the different notions of polarity. Obviously, further experimental work on these notions is required. In particular, the extent to which these notions overlap with each other and the extent to which they influence ITA strength remains unclear. Future experiments might specifically test for ITA effects with antonymic pairs in which the dimensional, the evaluative, and the markedness notion of polarity do not map onto each other, such as *old/young, heavy/light*—having in mind, however, that few adjectival pairs meet this constraint. Our approach also raises the issue whether it is sensible to maintain the dichotomous distinction positive versus negative or whether it makes more sense to continue with a continuous positivity notion. Although for the sake of clarity we did not go into this, our results also suggest that antonymic pairs differ in the extent to which they display ITA and polarity differences.

Abbreviations

NEG = negation, FEM = feminine, ITA = inference towards the antonym, COCA = the corpus of contemporary American English, NACH = negative adjectives complexity hypothesis

Additional File

The additional file for this article can be found as follows:

- **Appendix 1.** Average results on the two inference tests for positive and negative members of antonymic pairs (means, standard deviations, and results of the Wilcoxon signed rank tests; the pairs in bold were not used in the reported analyses). We used Wilcoxon signed rank tests to determine whether the inference scores for the two members of each of the antonymic pairs differ significantly. DOI: https://doi.org/10.5334/gjgl.151.s1

Competing Interests

The authors have no competing interests to declare.

References

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas M. & Deepayan Sarkar. 2007. Lme4: Linear mixed-effects models using S4 classes. *R package version*, 0.99875–6.

Becker, Israela. 2015. The good, the not good, and the not beautiful: On the non-obligatoriness of suppression following negation. *Corpus Linguistics and Linguistic Theory* 11(2). 255–283. DOI: https://doi.org/10.1515/cllt-2014-0010

Bierwisch, Manfred. 1989. The semantics of gradation. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives*, 71–261. Berlin: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-642-74351-1_3

Bolinger, Dwight. 1972. *Degree words*. Paris & The Hague: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110877786

Boucher, Jerry & Charles E. Osgood. 1969. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behaviour* 8. 1–8. DOI: https://doi.org/10.1016/S0022-5371(69)80002-2

Brasoveanu, Adrian & Jessica Rett. 2016. Evaluativity across adjective and construction types: An experimental study. *Journal of Linguistics*.

Brown, Penelope & Stephen Levinson. 1978. *Politeness: Some universals in language usage*. Cambridge & New York: Cambridge University Press.

Büring, Daniel. 2007a. More or less. *Proceedings of the Chicago Linguistic Society (CLS)* 43(2). 3–17.

Büring, Daniel. 2007b. Cross-polar nomalies. *Proceedings of Semantics and Linguistic Theory (SALT)* 17. 37–52. DOI: https://doi.org/10.3765/salt.v17i0.2957

Colston, Herbert. 1999. "Not good" is "bad", but "not bad" is not "good": An analysis of three accounts of negation asymmetry. *Discourse Processes* 28(3). 237–256. DOI: https://doi.org/10.1080/01638539909545083

Cruse, David A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Davies, Mark. 2008. *The corpus of contemporary American English: 520 million words, 1990-present*. Available online at: http://corpus.byu.edu/coca/.

Fillenbaum, Samuel. 1966. Memory for gist. Some relevant variables. *Language and Speech* 9. 217–227. DOI: https://doi.org/10.1177/002383096600900403

Fine, Kit. 1975. Vagueness, truth and logic. *Synthese* 30(3). 265–300. DOI: https://doi.org/10.1007/BF00485047

Fraenkel, Tamar & Yaacov Schul. 2008. The meaning of negated adjectives. *Intercultural Pragmatics* 5(4). 517–540. DOI: https://doi.org/10.1515/IPRG.2008.025

Giora, Rachel, Noga Balaban, Ofer Fein & Inbar Alkabets. 2005. Negation as positivity in disguise. In Herbert Colston & Albert Katz (eds.), *Figurative language comprehension: Social and cultural influences*, 233–258. Hillsdale, NJ: Erlbaum.

Grice, Henry P. 1975. Logic and conversation. In Peter Cole & Jerry H. Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. New York: Academic Press.

Heim, Irene. 1990. Presupposition projection. In Rob van der Sandt (ed.), *Presupposition, lexical meaning and discourse processes: Workshop Reader*. Nijmegen: University of Nijmegen.

Heim, Irene. 2006. Little. *Proceedings of Semantics and Linguistic Theory (SALT)* 16. 35–58. DOI: https://doi.org/10.3765/salt.v16i0.2941

Heim, Irene. 2008. Decomposing antonyms? In Atle Grønn (ed.), *Proceedings of Sinn und Bedeutung (SuB)* 12, 212–225. Oslo: ILOS.

Horn, Laurence R. 1984. Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Georgetown University round table on languages and linguistics*, 11–42. Washington DC: Georgetown University Press.

Horn, Laurence R. 1989. *A natural history of negation*. Chicago, IL: University of Chicago Press.

Israel, Michael. 2004. The pragmatics of polarity. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 701–723. Oxford: Blackwell.

Jespersen, Otto. 1924. *The philosophy of grammar*. London: Allen & Unwin.

Jones, Steven, Lynne Murphy, Carita Paradis & Caroline Willners. 2012. *Antonyms in English: Construals constructions and canonicity*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139032384

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical Linguistics* 1. 181–194. DOI: https://doi.org/10.1515/thli.1974.1.1-3.181

Kennedy, Christopher. 1997. *Projecting the adjective: The syntax and semantics of gradability and comparison*. Santa Cruz, CA: University of California Santa Cruz dissertation.

Kennedy, Christopher. 2007. Vagueness and grammar: The study of relative and absolute gradable predicates. *Linguistics and Philosophy* 30. 1–45. DOI: https://doi.org/10.1007/s10988-006-9008-0

Kennedy, Christopher & Louise McNally. 2005. Scale structure and the semantic typology of gradable predicates. *Language* 81(2). 345–381. DOI: https://doi.org/10.1353/lan.2005.0071

Krifka, Manfred. 2007. Negated antonyms: Creating and filling the gap. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 163–177. Houndmills: Palgrave Macmillan.

Kripke, Saul. 2009. Presupposition and anaphora: Remarks on the formulation of the projection problem. *Linguistic Inquiry* 40(3). 367–386. DOI: https://doi.org/10.1162/ling.2009.40.3.367

Lehrer, Adrienne. 1985. Markedness and antonymy. *Journal of Linguistics* 21. 397–429. DOI: https://doi.org/10.1017/S002222670001032X

Lehrer, Adrienne & Keith Lehrer. 1982. Antonymy. *Linguistics and Philosophy* 5(4). 483–501. DOI: https://doi.org/10.1007/BF00355584

Paradis, Carita & Caroline Willners. 2006. Antonymy and negation—The boundedness hypothesis. *Journal of Pragmatics* 38(7). 1051–1080. DOI: https://doi.org/10.1016/j.pragma.2005.11.009

Paradis, Carita, Joost van de Weijer, Caroline Willners & Magnus Lindgren. 2012. Evaluative polarity of antonyms. *Lingue e Linguaggio* 11(2). 199–214. DOI: https://doi.org/10.1418/38786

Reardon, James & Chip Miller. 2012. The effect of response scale type on cross-cultural construct measures: An empirical example using Hall's concept of context. *International Marketing Review* 29(1). 24–53. DOI: https://doi.org/10.1108/02651331211201534

Rett, Jessica. 2015. *The semantics of evaluativity*. Oxford: Oxford University Press.

Roche, Lisa. 2012. *La négation des adjectifs*. Paris: Université Paris 8 Master dissertation.

Sassoon, Galit W. 2012. The double nature of negative antonymy. In *Proceedings of Sinn und Bedeutung (SuB)* 16. 543–556.

Sassoon, Galit W. 2013. A typology of multidimensional adjectives. *Journal of Semantics* 30(3). 335–380. DOI: https://doi.org/10.1093/jos/ffs012

Singh, Raj, Evelina Fedorenko, Kyle Mahowald & Edward Gibson. 2016. Accommodating presuppositions is inappropriate in implausible contexts. *Cognitive Science* 40. 607–634. DOI: https://doi.org/10.1111/cogs.12260

Solt, Stephanie. 2015. Measurement scales in natural language. *Language and Linguistics Compass* 9. 14–32. DOI: https://doi.org/10.1111/lnc3.12101

van der Sandt, Rob & Bart Geurts. 2001. Too. In *Proceedings of the 13th Amsterdam Colloquium.* University of Amsterdam.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1–2). 1–77. DOI: https://doi.org/10.1093/jos/3.1-2.1

von Stechow, Arnim. 2009. The temporal adjectives früh(er)/spät(er) and the semantics of the positive. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness, and nominalization,* 214–233. Oxford: Oxford University Press.

Williamson, Timothy. 1994. *Vagueness.* London: Routledge.

Winterstein, Grégoire. 2011. The meaning of the additive "too": Presupposition and discourse similarity. In Nick Bezhanishvili, Sebastian Löbner, Kerstin Schwabe & Luca Spada (eds.), *Logic, language and computation. 8th international Tbilisi symposium on logic, language, and computation. Lecture notes in artificial intelligence* (LN6618), 322–341. Berlin: Springer.