

Israela Becker

## The good, the not good, and the not beautiful: On the non-obligatoriness of suppression following negation

**Abstract:** The view that suppression of a concept within the scope of negation is not unconditional was originally introduced by Giora (2003, 2006; Giora and Fein 1999; Giora et al. 2007) via the *retention hypothesis*. Giora and her colleagues argue that negation does not necessarily suppress the concept within its scope. Instead, it often *retains* it for pragmatic considerations, both in the mind of the speaker and the addressee. The present study provides a quantitative corpus-based test for the retention hypothesis, that is the non-obligatoriness of suppression of negated concepts (also known as the *negation as mitigation* hypothesis, Giora 2003; Giora et al. 2005b), via a two-pronged method which combines corpus data and behavioral data. It focuses on the notion of *polarity strength*, which is a numerical value disclosing the degree of positivity or negativity associated with an adjective. A simple statistic which is introduced for the sake of this study – the *Strength Index (SI)* – naïvely assumes that canonical adjectives can be mitigated by replacing them with their negated antonyms, thus making it possible to attribute SI to them. SI is calculated for 8 canonical adjectival antonymous pairs of an emotive nature (such as *good-bad*). Depending on prior positive expectations, the retention hypothesis will gain support if the following results are obtained: Correlation between the SIs of unfavorable adjectives (e.g., *bad*) and behavioral data on the one hand, and the lack of correlation between the SIs of favorable adjectives (e.g., *good*) and behavioral data, on the other hand. Results attest to this correlation pattern, providing support for the retention hypothesis (see also Colston 1999).

**Keywords:** psycholinguistics, negation, mitigation, retention, suppression, polarity strength, customer reviews

DOI 10.1515/cllt-2014-0010

# 1 Introduction

According to the *negation as suppression* view, negation is regarded as an instruction from a speaker to an addressee to eliminate the concept within the scope of negation from the mental representation and replace it with its antonym, if available. If no antonym is available, then the negated concept simply decays. (For a review of the suppression hypothesis see Giora et al. 2007, who critically reviewed prominent works in this field: Hasson and Glucksberg 2006; Kaup et al. 2006; MacDonald and Just 1989; Mayo et al. 2004; and others).

A different, pragmatic approach to negation known as *negation as mitigation* was proposed and tested by Giora and colleagues (Giora 2006; Giora and Fein 1999; Giora et al. 2005a, 2005b, 2007) and briefly discussed earlier (with respect to morphological negation) by other linguists (Horn 1989: 273; Jespersen 1917: 36). The mitigation hypothesis assumes that negation is *not* necessarily a suppressor, reducing the activation levels of the negated concept and replacing it with its alternative opposite (if available). Instead, this hypothesis postulates that the negation operator is sensitive to discourse goals. Often, rather than suppressing it, the negation operator *retains* the concept within its scope while mitigating it. Thus, *not fresh* is less than *fresh*, whereas *not stale* is less than *stale* but not quite *fresh*. Giora and colleagues further showed that when ongoing discourse is relevant to prior discourse in which a concept is negated, that concept is retained in memory rather than suppressed. Results suggest that comprehenders are more sensitive to discourse goals than to negation (Giora et al. 2007).

The present research tests the negation as mitigation hypothesis, i.e., the conditional retention of the concept in the scope of negation, via a two-pronged method, which combines corpus data with behavioral data.

It should be noted, however, that corpus studies cannot provide direct evidence with regard to online processes, such as suppression or retention of concepts. They can only be supportive (or not) of predictions in this respect. However, the results presented below are consistent with only one of the two options, namely, retention of negated concepts, and are similar to the results of offline experiments presented in Giora et al. (2005b).

## 1.1 On negation as mitigation

As noted above, Giora et al. (2005a) were the first to provide experimental evidence that the negation operator mitigates rather than suppresses the mental representation of the negated concept. They showed that both “X is Y” and “X is not Y” (e.g., “This instrument is *sharp*” and “This instrument is not *sharp*”)

equally prime a concept (*piercing*) related to the target Y (e.g., *sharp*), regardless of negation; in contrast, the antonymic prime of the target Y (e.g., “This instrument is *blunt*”) did not. The fact that negation does not have an initial effect on the concept within its scope, they argued, must affect the representation of the negated concept. To do that, they collected scalar ratings of favorable (e.g., *sharp*) and unfavorable (e.g., *blunt*) concepts of antonymous pairs either modified by a negative marker (*no*), an affirmative marker (*fairly*), or not modified at all. Results showed that the negation operator did not suppress the negated concept but only mitigated it. Thus, on a scale ranging between *sharp* (at the positive end of the scale) and *blunt* (at the negative end of the scale), *not sharp* was rated as significantly more favorable than *blunt*, while *not blunt* was rated as significantly more unfavorable than *sharp* and less unfavorable than *fairly blunt*. Mitigation effects were similar for both favorable (e.g., *sharp*) and unfavorable (e.g., *blunt*) concepts.

Giora et al. (2005b) further showed that speakers are aware of this mitigation effect of the negation operator. In their study, participants were presented with sentences such as “What you said was a lie” and “What you said was not true”. They were asked to decide which of the two sentences they would use if they wished to be polite. Results show that the participants favored the negated version.

Results of Colston (1999), Fraenkel and Schul (2008), and Paradis and Willners (2006) are in agreement with Giora’s notion of mitigation, although some of them (Colston as well as Fraenkel and Schul) did not explicitly state that in order for mitigation to take place, the concept within the scope of negation must be retained rather than suppressed. However, contrary to Giora’s findings, they showed that the mitigating effect of negation is not necessarily symmetric; that is, it depends on the polarity of the negated concept, on speaker’s prior expectations, and on the nature of the adjectives, whether contrary (such as *industrious-lazy*) or contradictory (such as *alive-dead*).

Colston (1999) presented participants with written scenarios which were intended to evoke either a positive or a negative expectation. Each scenario was concluded with a comment made by one of the characters in the scenario. The comment contained a favorable adjective like *good*, an unfavorable adjective like *bad*, or a negated version of an adjective like *not good* or *not bad*. Participants were asked to rate the adequacy of the concluding comment vis à vis the preceding scenario. Under positive expectations, Colston observed that the use of a negated favorable adjective (e.g., *not good*) is *equivalent* to the use of an unfavorable adjective (e.g., *bad*), whereas a negated unfavorable adjective (e.g., *not bad*) could not replace its favorable antonym (e.g., *good*). Under negative expectations, this asymmetry was not obtained. Colston explained his results by integrating

politeness theory (Brown and Levinson 1987) into relevance theory (Sperber and Wilson 1986/1995); more specifically, by assuming that verbal politeness is a contextual assumption of relevance theory.

Whereas Colston focused on the effect of expectations on mitigation, Paradis and Willners (2006) examined the sensitivity of negation to the type of antonymy, whether contrary or contradictory. Unlike Colston, who studied adjectival pairs of an emotive nature (e.g., *happy-sad*), Paradis and Willners focused on adjectival pairs of a neutral character (e.g., *wide-narrow*) aiming to determine the extent to which an adjective is similar to its negated antonym. They claimed that the interaction of the negation operator with a contrary adjective retains the negated item, whereas the interaction of the negation operator with a contradictory adjective suppresses it (and replaces it with its antonym). Paradis and Willners asked participants to rate individual adjectives on an 11-point scale: Participants were shown a sentence such as “The road along the coast is not wide” and were asked to decide whether the road is more like a path (which represented the lower end of the 11-point scale) or a more like motorway (which represented the upper end of the 11-point scale). It is not only that Paradis and Willners’ findings supported their hypothesis, they also showed that the antonymic relation (i.e., contrary or contradictory) is not necessarily symmetric: while *fertile* is contrary to *sterile*, *sterile* is practically contradictory to *fertile*.

Like Paradis and Willners, Fraenkel and Schul (2008) were first and foremost concerned with the effect of the nature of the antonymous adjectival pair – whether contrary or contradictory – on mitigation. They presented participants with pairs of sentences: one sentence was affirmative, for instance, “The coffee is hot”, and the other was the negated version of the previous sentence – “The coffee is not cold”. Participants rated the resemblance of the two sentences in each pair on a 20-point scale ranging from *completely different* to *completely identical*. Fraenkel and Schul showed that negation of an adjective in a contrary pair (e.g., *smooth-rough*) is more likely to affect mitigation of the antonymic concept than negation of an adjective in a contradictory pair (e.g., *together-apart*). Moreover, the mitigation effect was shown to depend on the type of the negated adjective: The negation of an unfavorable adjective (e.g., *not bad*) results in a stronger mitigation of its (favorable) counterpart (e.g., *good*). In contrast, negating the favorable adjective (e.g., *not good*) results in a weaker mitigation of its (unfavorable) counterpart (e.g., *bad*). I.e., *not good* resembles *bad* to a greater extent than *not bad* resembles *good*. Unlike Paradis and Willners, but like Colston, the adjectives examined by Fraenkel and Schul were of an emotive nature.

In addition, Fraenkel and Schul showed that the effect of negation on mitigation has also to do with prior expectations (as previously shown by Colston): When examined in a context in which participants’ expectations were deliber-

ately raised, the extent of mitigation of unfavorable adjectives (e.g., *bad*) in negative contexts proved to be identical to the extent of mitigation of favorable adjectives (e.g., *good*) in positive contexts.

All in all, the aforementioned studies show that retention of a concept in the scope of negation depends on the nature of the negated adjectives, whether contrary or contradictory, as well as on interlocutors' prior expectations and discourse goals.<sup>1</sup>

## 1.2 Evidence from production

Along the same lines of the aforementioned studies, I also examine the relation between an adjective *X* and its negated antonym, *not Y*. But unlike them, I examine this relation via corpus while the aforementioned studies examined this relation based on participants' judgments.

It should be noted that suppression and retention are regularly associated with comprehension rather than with production of utterances. And indeed, all the aforementioned studies examined how addressees comprehend utterances. I, on the other hand, examine suppression and retention in the context of producing utterances. In doing so, I draw on psycholinguists such as Giora (2003), Levelt (1989), Stephens et al. (2010), and Pickering and Garrod (2013; and see references therein), just to mention a few who regard production processes as mirroring comprehension processes. This unified view, according to which what the speaker says is based upon her prediction of how the addressee processes incoming messages, licenses the analysis of retention and suppression of concepts in the scope of negation with respect to production rather than only (traditionally) with respect to comprehension.

And again, it should be noted that corpus studies cannot provide direct evidence with regard to online processes, such as suppression or retention of

---

<sup>1</sup> Retention – rather than automatic suppression – of concepts in the scope of negation was also demonstrated by Autry and Levine (2012, 2014) and by Shuval and Hemforth (2008), who followed a somewhat different line of research: While the works examined here focused on retention of *adjectives* in the scope of negation, thus associating retention with mitigation, Autry and Levine as well as Shuval and Hemforth focused on retention of *entities* in the scope of negation, where mitigation is no longer relevant. Autry and Levine measured the activation levels of unlicensed negated entities as opposed to licensed negated entities; Shuval and Hemforth focused on the way entities were negated, comparing ordinary negation with repair-like negation. Both concluded, consistently with Giora's (2003, 2006) functional approach to negation, that entities in the scope of negation are retained as a consequence of discourse requirements.

concepts. They can only be supportive (or not) of predictions in this respect. However, the results presented below are consistent with only one of the two options, namely, retention of negated concepts, and are similar to the results of offline experiments presented in Giora et al. (2005b).

This article is organized as follows: I first present the well-known concept of *polarity strength* and subsequently provide my new formulation to it, the *Strength Index*, which is the tool used for testing the retention hypothesis. I, then, lay out the retention hypothesis and the suppression hypothesis, and make predications about the correlation pattern between speakers' evaluation of the strength of adjectives (i.e., behavioral data) and the corpus-based Strength Index results if retention of the concept in the scope of negation takes place or if suppression of the concept in the scope of negation takes place. Next, the collected data are presented in three phases: First, The appropriateness of the corpus with regard to prior discourse expectations is examined; second, the procedure of choosing the appropriate adjectives, for which the Strength Index is calculated, is described; finally, participants' ratings of the strength of the adjectives – for which the Strength Index has been calculated – are presented, and the correlation pattern between these ratings and the Strength Index results is analyzed in view of the predictions of the hypothesis tested here.

## 2 Polarity strength

*Polarity strength* is a numerical value disclosing the degree of positivity or negativity associated with a concept, i.e., the strength of the sentiment expressed when using this concept. Polarity strength is often referred to as potency or intensity.

Attempts to determine polarity strength date back to the late 1950s when Osgood et al. (1957) developed a technique for measuring the connotative meaning of concepts, known as the *semantic differential*. A semantic differential scale is a scale displaying two bipolar adjectives (i.e., morphologically unrelated concepts that define the two ends of an evaluative scale) such as *good-bad*, *active-passive* or *strong-weak*. Osgood and colleagues asked participants to express their attitude towards concepts (of which some were adjectives) along many semantic differential scales. Their extensive work, as well as the work of Jenkins (1960) and Jenkins et al. (1958) resulted in exhaustive lists of concepts rated along many dimensions of which the *potency scale* appeared to be one of the most important scales that participants used to evaluate a given concept. The rating of a concept along the potency scale defines, in fact, the polarity strength of the concept.

Inspired by Osgood et al. (1957), Kamps et al. (2004) tried to calculate the polarity strength of concepts (adjectives only) by using the WordNet database

(Miller et al. 1990).<sup>2</sup> Kamps et al. measured the shortest distances (“minimal path lengths”) between a concept and each of the two ends of the scale defined by the two antonyms *strong-weak* (the potency scale, according to Osgood et al.) in terms of nodes (which are WordNet words). The difference between these two shortest distances was normalized to the distance between *weak* and *strong* so as to yield a value within the interval  $[-1,+1]$ . The value obtained was the *potency factor*, which is, in fact, what is referred to here as polarity strength. The more positive the potency factor, the stronger the adjective. The more negative the potency factor, the weaker the adjective.<sup>3,4</sup>

In this article, I propose a different potency measurement: To illustrate, let us, for the moment, *naïvely* assume that an adjective can be mitigated by its negated antonym (e.g., *bad* and *not good*). When a speaker intends to use a certain concept, she can either use the concept (e.g., *bad*) or she can use its negated equivalent (e.g., *not good*) for the sake of mitigating the concept.

The substitution of an adjective for a single negated antonym rather than other (negated or non-negated) options can hold for canonical adjectival pairs, which are defined by Jones et al. (2007) as such “if the two words are associated by ‘convention’ as well as by semantic relatedness, for example *public/private*. In other words, canonical antonym pairings have been learnt as pairings of lexical units not just derived by semantic rules” (p. 131).

In other words, if canonical adjectival pairs are at focus, then one can assume that if a concept (e.g., *bad*) is avoided, it must be replaced by its negated antonym (e.g., *not good*).

Statistically speaking, a canonical adjective and its negated antonym are the two possible outcomes that constitute a sample space, which allows the speaker to choose either one or the other. Each of these two outcomes could be assigned a probability. The sum of the two probabilities amounts to 1.

It has been demonstrated that addressees are prone to accept a message if they regard the speaker as a credible source of information (for a critical review on source credibility as a major factor in message acceptance see Pornpitakpan 2004; and references therein). Addressees assess speaker’s credibility based on their prior beliefs regarding the speaker, on the speaker’s credentials, and on the message itself (Slater and Rouner 1996). A speaker the addressees know nothing about, i.e., a speaker of a-priori low credibility, would benefit from using low

---

<sup>2</sup> <http://wordnet.princeton.edu/>

<sup>3</sup> In fact, Kamps et al. extended this procedure, using two additional scales: the evaluative scale (*good-bad*) and the activity scale (*active-passive*).

<sup>4</sup> Other works that attempt to computationally determine the polarity strength of single concepts are Esuli and Sebastini 2006; Godbole et al. 2007; and Williams and Anand 2009.

intensity rather than high intensity language to enhance message impact (e.g., Bradac et al. 1979; Franzwa 1969; and Hamilton et al. 1990). This should be the case, for example, with authors of online reviews who are often anonymous to the addressees and, therefore, of a-priori low credibility. Authors of online reviews, who wish to get their message across, are expected to refrain from using intense language (Jensen et al. 2013). Consequently, the stronger an adjective, the more infrequent it should be compared to its mitigated version which is – according to the mitigation/retention hypothesis – the negated antonym of the adjective. And so a *Strength Index* (*SI*) can be formulated:

$$SI_{\text{Adjective}} = \frac{(\text{Negated\_Antonym})}{(\text{Negated\_Antonym} + \text{Adjective})}$$

where *SI* stands for *Strength Index*; The term *Adjective* refers to the number of times an adjective appears in a corpus (*bad*, for instance); the term *Negated Antonym* refers to the number of times the negated antonym (e.g., *not good*) appears in the same corpus. Note that the denominator of the *SI* expresses the availability of a concept and its negated antonym. The ratio between the numerator (which is the negated option) and the denominator (which is the sum of the adjective and its negated antonym) expresses the extent to which a negated adjective is preferred over its antonym. The higher the *SI*, the less preferred (and consequently, more often replaced) the adjective is with respect to its negated antonym; that is, the higher the *SI*, the stronger the adjective.<sup>5</sup>

### 3 Hypothesis and predictions

According to the retention hypothesis, suppression is not obligatory but sensitive to discourse goals. Hence if *positive* prior expectations are frustrated, *not good* rather than *bad* is the preferred option; if *negative* prior expectations are thwarted then *not bad* rather than *good* is the preferred option. In both cases, the negated version mitigates its affirmative counterpart, allowing the concept in the scope of negation to resonate with frustrated expectations.

---

<sup>5</sup> Note that the potency measurement that I propose here is much more limited than the ones proposed by other researches (Esuli and Sebastiani 2006; Godbole et al. 2007; Kamps et al. 2004; Williams and Anand 2009), whose methods can be used to calculate the polarity strength of each and every adjective. My method, however, is adequate for canonical adjectival pairs only. Yet, one must bear in mind that while other researchers aim at providing a practical tool to use for automatic sentiment classification of large chunks of texts, I aim at testing the non-obligatory suppression of a concept within the scope of negation.

Here I will be testing this hypothesis by looking at the correlation (or the lack of it) of the numerical values of the SI attributed to adjectives with speakers' evaluation of the strength of these adjectives, given that expectations are positive.

Basically, the validity of SI as a measure of potency of an adjective draws on the relation between an adjective and its negated antonym: the SI expresses the strength of an adjective only if its negated antonym serves as the mitigated version of the adjective. Specifically, correlation between speakers' evaluation of strengths of adjectives and SI values indicates that the negated antonym is indeed the mitigated version of the adjective, so that the formulation of SI correctly captures the relation between the adjective and its negated antonym. Lack of correlation between speakers' evaluation of strengths of adjectives and SI values indicates that the negated antonym is not the mitigated version of the adjective, and that the formulation of SI fails to capture the relation between the adjective and its negated antonym. Either correlation or the lack of correlation can be accounted for by retention or suppression of the concept in the scope of negation. Naturally, the decision whether retention or suppression takes place must conform with prior discourse expectations.

When prior positive expectations are assumed, 4 correlation patterns are possible, as summarized in Table 1. Pattern 1 and pattern 2 are detailed below (where *good*, *not good*, *bad* and *not bad*, are used to exemplify the various possible combinations of favorable and unfavorable adjectival pairs). The other two patterns can be analyzed along the same lines.

*Pattern 1.* Following Colston (1999), positive expectations make *good* a natural candidate for a speaker to use. *Bad* as well as *not bad* are discordant with positive expectations. If the concept in the scope of negation is retained, as follows from the retention hypothesis, then *not bad* will not replace *good*, because it would conflict with prior positive expectations. Therefore, *good* and *not bad* will not constitute a single sample space. Consequently, the SI of favorable adjectives (e.g., *good*) will fail to express the strength of favorable adjectives (because *good* cannot be replaced by *not bad*), and speakers' evaluation of the strength of favorable adjectives will not be correlated with SI values.

If the concept in the scope of negation is indeed retained, but positive expectations are frustrated, then *bad* (which is not part of positive expectations) is expected to be replaced by *not good*, as *good* should be the natural candidate when it comes to positive expectations. Therefore, *bad* and *not good* will make up a single sample space, which allows *not good* to replace *bad*. Consequently, the SI of unfavorable adjectives (e.g., *bad*) will correctly capture the strength of unfavorable adjectives, and speakers' evaluation of the strength of unfavorable adjectives will be correlated with SI values.

In sum, the only way for such correlation pattern to obtain – correlation for unfavorable adjectives (e.g., *bad*) but lack of correlation for favorable adjectives

(e.g., *good*) – is when the concept in the scope of negation is retained and can, therefore, resonate with prior expectations. If this correlation pattern is found, this will support the prediction that the concept within the scope of negation is retained rather than suppressed.

*Pattern 2.* If speakers' evaluation of the strength of adjectives is in correlation with SI values of both favorable and unfavorable adjectives, then *not good* is the mitigated version of *bad* and *not bad* is the mitigated version of *good*. In other words, *bad* and *not good* make up a single sample space; *good* and *not bad* also make up a single sample space. Assuming prior positive expectation, *bad* is expected to be replaced by *not good* and this can be accounted for by the retention of the negated concept (*good*), which is congruent with positive expectations (as in pattern 1). However, since positive expectations do not take *bad* into account, *not bad* is in conflict with prior positive expectations, and *not bad* is not the mitigated version of *good* unless suppression of *bad* in the scope of negation is assumed. In sum, this correlation pattern – correlation for unfavorable adjectives (e.g., *bad*) and correlation for favorable adjectives (e.g., *good*) too – is congruent with prior positive expectations only if favorable adjectives in the scope of negation (e.g., *good* in the scope of *not good*) are retained, while unfavorable adjectives in the scope of negation (e.g., *bad* in the scope of *not bad*) are suppressed and replaced by their favorable antonyms (*good*). Such findings will only partially support the retention hypothesis.

Pattern 3 and pattern 4, which are listed in Table 1, can be analyzed along the same lines. The only correlation pattern, then, that fully supports the retention

**Table 1:** The potential correlation patterns between participants' evaluation of the strength of favorable and unfavorable adjectives and SI values (given prior positive expectations)

|           | <b>Correlation between participants' evaluation and SI values of:</b> | <b>Lack of correlation between participants' evaluation and SI values of:</b> | <b>Retention or suppression? (when prior positive expectations are assumed)</b> |
|-----------|---|---|---|
| Pattern 1 | Unfavorable adjectives  | Favorable adjectives  | Retention<br>Retention  |
| Pattern 2 | Unfavorable adjectives<br>Favorable adjectives                        |   | Retention<br>Suppression  |
| Pattern 3 |   | Unfavorable adjectives<br>Favorable adjectives                                | Suppression<br>Retention  |
| Pattern 4 | Favorable adjectives  | Unfavorable adjectives  | Suppression<br>Suppression  |

hypothesis – given prior positive expectations – is pattern 1, which shows correlation between speakers' evaluation of the strength of unfavorable adjectives and their SI values, and at the same time, lack of correlation between speakers' evaluation of the strength of favorable adjectives and their SI values.

## 4 Corpus characterization

The corpus used to test the retention hypothesis is the Blitzer et al. (2007) Sentiment Dataset<sup>6</sup> which focuses on customer reviews taken from the amazon.com website.<sup>7</sup>

Taboada (2011) analyzes online reviews (focusing on online movie reviews) showing that this genre is both descriptive and evaluative. De Jong and Burgers (2013) show that consumer reviews are more evaluative than descriptive (as compared to professional critics). Assuming that the amazon.com review system relies largely on consumer reviews rather than on professional critics, this corpus is likely to provide a large pool of adjectives for my analysis.

The corpus consists of 7,854 customer reviews in which each review includes a customer's rating of a product on a 1- to 5-star scale. The corpus covers four product types (domains): Books, DVDs, Kitchenware, and Electronics. Each domain contains ~2,000 reviews, of which about half are positive, and half are negative. The corpus consists of ~1,100,000 tokens, of which ~350,000 are equally divided between positive and negative book reviews, ~350,000 are equally divided between positive and negative DVD reviews, ~190,000 are equally divided between positive and negative kitchenware reviews, and ~220,000 are equally divided between positive and negative consumer electronics reviews. The reviews were written by 7,192 different reviewers (distinguishable by their names). Each reviewer reviewed products in only one category of the four categories listed above. Of the 7,192 reviewers, 95.08% wrote one review only, 4.10% wrote two reviews, and the remaining ~1% wrote more than two reviews.

This corpus is expected to be a corpus in which the authors of the reviews are such that have positive expectations, because customers of consumer products naturally expect the products that they have purchased to be (at least) satisfactory. The following examples are examples that faithfully represent the spirit of this corpus, where reviewers express their prior positive expectations whether they write a positive or a negative review:

---

6 <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html>

7 <http://www.amazon.com>

- (1) We would expect a well-balanced approach to writing about the work of the Holy Spirit in the church by the author who also wrote “Tyranny of the Urgent” and we get it! (a positive book review authored by “R. L. Stites”)
- (2) I usually read Stephen King for his interesting ideas and imagery and expect his books to read fast and well. (a negative book review authored by “Filmbuff-reads stuff”)
- (3) I expect to enjoy my new Gold Dust set for at least, the next 20 years. (a positive review of a dinnerware set authored by “Mrs. Denise M. Siracusa”)
- (4) It looks OK, but it just isn’t the quality item you would expect with the name Cuisinart. (a negative review of a rice cooker authored by “D. ERMER AtoZ Junkie”)

In this article I focus on adjectives on which there is a consensus regarding their polarity. Such adjectives can be found in two of the lists of the modern version of the *General Inquirer*<sup>8</sup> (Stone et al. 1966): One list contains positively evaluated nouns, verbs, adjectives, and adverbs, and the other contains negatively evaluated nouns, verbs, adjectives, and adverbs (altogether 1,524 adjectives, 1,416 nouns, 1,073 verbs and 187 adverbs). I narrowed down these two lists into adjectival lists only, from which I further extracted items that would occupy ends of bipolar scales, while removing items that would occupy ends of unipolar scales (i.e., adjectives which are morphologically related antonyms such as *responsible* and *irresponsible*), to form a pool of potential canonical antonymous pairs (as indicated in Section 2). Eventually, a list of 700 favorable adjectives (e.g., *good*) and 623 unfavorable adjectives (e.g., *bad*) was selected.

I then counted the total number of occurrences of logically equivalent counterparts (adjectives from the above lists and their negated antonyms) in reviews of both polarities: (i) the total number of favorable adjectives and their negated unfavorable counterparts (*clean* and *not dirty*, for instance), and (ii) the total number of unfavorable adjectives and their negated favorable counterparts (*dirty* and *not clean*, for example). The negation operators used in the counting procedure were all function words: *no*, *not*, *never*, *none*, *neither*, *nor*, *nowhere*, *nobody*, *nothing*, *-n’t* as listed in Tottie (1991) and in Tottie and Paradis (1982).<sup>9</sup>

Summing across all review categories, there are altogether 35,801 occurrences of favorable and negated unfavorable adjectives (e.g., *good* and *not bad*, respectively) while only 20,159 occurrences of unfavorable and negated favorable adjectives.

<sup>8</sup> <http://www.wjh.harvard.edu/~inquirer/>

<sup>9</sup> I am aware of negating content words such as *fail*, *lack* or *avoid*. However, these negation terms are quite infrequent in the corpus relative to the negating function words which are listed in the text (Tottie 1991; Tottie and Paradis 1982), and thus have a minor effect on the calculations.

tives (e.g., *bad* and *not good*, respectively). It seems that reviewers are positively biased: they are more likely to praise a product than to criticize it.

## 5 Strength index calculations

Deciding upon the adjectival pairs to test the retention hypothesis is a task that requires the specification of several criteria, as follows:

1. As already noted, adjectival pairs must be bipolar, rather than unipolar. Removal of unipolar adjectival pairs from the adjectival lists is meant to avoid morphologically embedded negation, which provides no other option but to use a negated adjective rather than an affirmative alternative.
2. Adjectival pairs must also constitute canonical or near-canonical pairs in order to delimit the sample space from which the speaker is to choose the adjective to be used (see Section 2); the more optional antonyms exist for a single adjective, the more the counts (and consequently the calculations) are susceptible to inaccuracies.
3. For the sake of reliable statistics, selected adjectival pairs must appear in the corpus frequently. Such pairs should include common adjectives. Common adjectives are defined by Deese (1965) “as those which occur with a frequency of fifty instances per million words or more in the Thorndike-Lorge count” (p. 121). For the sake of the current calculation, I relaxed this criterion so that each of the groups { unfavorable adjective, negated favorable antonym } and { favorable adjective, negated unfavorable antonym } must occur at a frequency of fifty instances (or more) per million words.
4. Adjectives must be *predicative adjectives*, that is, “adjectives that can be used as predicates in sentences of the form *NP is Adj*” (Gross et al. 1989: 93), since *non-predicative adjectives* (i.e., attributive adjectives) of the form *Adj NP* are regarded as stylistic variants of modifying nouns and, as such, do not necessarily have antonymous counterparts (Levi 1978). The non-predicative version of a predicative adjective is included in the counting procedure, provided that both versions have the same meaning, such as *this man is rich/this is a rich man*, but unlike *this lawyer is criminal/ this is a criminal lawyer*.

Criterion 1 had already been fulfilled, as I used the filtered adjectival lists of the General Inquirer as explained in Section 4.

Criteria 2, 3, and 4, which were considered simultaneously, produced nine adjectival pairs that have been established as canonical by several research groups (Deese 1964; Gross et al. 1989; Jones et al. 2007; Paradis 2010; Van de Weijer et al. 2012). These nine canonical adjectival pairs and the references asserting their canonicity are listed in Table 2.

**Table 2:** The antonymous adjectival pairs that were used for the calculation of SIs: Canonical pairs are underlined. The direct synonyms of adjectives appear below the adjectives in the same box. Each adjective is accompanied by its salient (or relevant) definition according to WordNet, a representative example from the corpus, and references supporting the canonicity of the adjectival pair (refer to the text for further details)

| Pair | Unfavorable adjective            | Definition and example  | Favorable adjective        | Definition and example  | References for canonicity or near-canonicity              |
|------|----------------------------------|---|----------------------------|---|---|
| 1    | <u>Bad</u> :                     | Having undesirable or negative qualities<br>E.g., "I don't see why that is a <u>bad</u> thing."   | <u>Good</u> :              | Having desirable or positive qualities especially those suitable for a thing specified<br>E.g., "It has a lot of <u>good</u> stories that are very entertaining and each of them illuminates important spiritual truths."             | Deese 1964;<br>Paradis 2010;<br>Van de Weijer et al. 2012 |
| 2    | <u>Shallow</u> :<br>Superficial: | Lacking depth of intellect or knowledge<br>Not deep or penetrating emotionally or intellectually:<br>E.g., "The entire plot is self-serving, <u>shallow</u> , and unrealistic." | <u>Deep</u> :<br>Profound: | Marked by depth of thinking<br>Showing intellectual penetration or emotional depth<br>E.g., "If you are interested in <u>deep</u> philosophical writings on technology and the human condition, with links to literature, read this." | Deese 1964  |
| 3    | <u>Wrong</u> :                   | Not correct; not in conformity with fact or truth<br>E.g., "Have you ever seen someone else do the exact same <u>wrong</u> thing and condemn him?"                              | <u>Right</u> :             | Free from error; especially conforming to fact or truth<br>E.g., "The Bad Reviews Are <u>Right</u> – Don't Buy It!!!"   | Deese 1964; Van de Weijer et al. 2012                     |

Table 2 (cont.)

| Pair | Unfavorable adjective  | Definition and example   | Favorable adjective                 | Definition and example   | References for canonicity or near-canonicity                                    |
|------|------------------------|--|-------------------------------------|--|---|
| 4    | <u>Ugly</u> :          | Displeasing to the senses<br>E.g., "Looks great sitting on the kitchen counter, whereas many of the other toaster models today look like <u>ugly</u> chrome spaceships from the 1950's." | <u>Beautiful</u> :<br><br>Pretty:   | Delighting the senses or exciting intellectual or emotional admiration<br><br>Pleasing by delicacy or grace<br><br>E.g., "Otherwise, as usual with Macs, very easy to use, even for low-tech people like myself, <u>beautiful</u> design, small and light, not too fragile, just great on the go." | Deese 1964;<br>Gross et al. 1989;<br>Paradis 2010;<br>Van de Weijer et al. 2012 |
| 5    | <u>Boring</u><br>Dull: | So lacking in interest as to cause mental weariness<br><br>E.g., "The un witty dialogue tries to be clever all throughout the movie and is just sooo <u>boring</u> ."                    | <u>Interesting</u><br><br>Exciting: | Arousing or holding the attention<br><br>Stimulating interest and discussion<br><br>E.g., "The Orbit's special effects are <u>interesting</u> and fun."  | Paradis 2010;<br>Van de Weijer et al. 2012                                      |
| 6    | <u>Sad</u> :           | Experiencing or showing sorrow or unhappiness<br><br>E.g., "That was <u>shocking</u> and <u>sad</u> ."   | <u>Glad</u> :<br><br>Happy:         | Showing or causing joy and pleasure; especially made happy<br><br>Enjoying or showing or marked by joy or pleasure<br><br>E.g., "I was SO Glad she canceled the tour and I got my money back."   | Deese 1964;<br>Paradis 2010;<br>Van de Weijer et al. 2012                       |

Table 2 (cont.)

| Pair | Unfavorable adjective         | Definition and example   | Favorable adjective      | Definition and example  | References for canonicity or near-canonicity                                    |
|------|-------------------------------|--|--------------------------|---|---|
| 7    | Hard<br>Difficult:            | Not easy; requiring great physical or mental effort to accomplish or comprehend or endure<br>E.g., "And considering how <u>hard</u> its been finding these items, who can complain." | <u>Easy</u> :<br>Simple: | Posing no difficulty; requiring little effort<br>Easy and not involved or complicated<br>E.g., "Installation was super <u>easy</u> ."   | Deese 1964;<br>Gross et al. 1989;<br>Paradis 2010;<br>Van de Weijer et al. 2012 |
| 8    | <u>Expensive</u> :<br>Costly: | High in price or charging high prices<br>Having a high price<br>E.g., "You can just get better performance from a less <u>expensive</u> web cam."                                    | <u>Cheap</u> :           | Relatively low in price or charging low prices<br>E.g., "but considering the age of these movies and the <u>cheap</u> price tag this is an amazing deal!"   | Van de Weijer et al. 2012   |
| 9    | <u>Dirty</u> :                | Soiled or likely to soil with dirt or grime<br>E.g., "This printer was fine until I changed cartridges and a contact was <u>dirty</u> ."   | <u>Clean</u> :           | Free from dirt or impurities; or having clean habits<br>E.g., "I still give the case a 4 star rating, because I am a soldier in Iraq, and this case is doing a very good job of keeping my IPOD <u>clean</u> ." | Deese 1964;<br>Paradis 2010;<br>Van de Weijer et al. 2012                       |

Adjectives, like other lexemes, could have several meanings. For each of the 18 adjectives in the 9 canonical pairs, I focused on the globally salient meaning as it appears in WordNet<sup>10</sup>, to satisfy the requirement for large statistics. However, the globally salient meaning is not always of an emotive nature and therefore not positively or negatively biased; or it does not always produce good statistics in the corpus. For example: *shallow* has a concrete definition (“lacking physical depth”) and an abstract one (“lacking depth of intellect or knowledge”). The concrete definition has the globally salient meaning, but it is not of an emotive nature. Even if it was of an emotive nature, the corpus contains book and DVD reviews for which *shallow* in the abstract sense is, not surprisingly, more relevant than *shallow* in the concrete sense. Therefore, I considered the abstract meaning of *shallow* as the salient one.

Another adjective that deserves special attention is *cheap* which is considered favorable with respect to the price of a product, but is considered unfavorable with respect to product quality. I considered *cheap*'s favorable meaning (following the General Inquirer), and consequently *expensive* was considered its unfavorable antonym. The WordNet definitions of *shallow* and *cheap* as well as the definitions of the other 16 adjectives, are listed in Table 2.

For the sake of accurate quantification, one cannot settle for counting the occurrences of a single adjective, even though canonical adjectives are at focus. Rather, it is preferable to consider synonyms of the very same definition. I used the *direct* synonym of each member of the nine adjectival pairs as suggested by WordNet. The direct synonym of each adjective sets the boundaries for each sample space. The direct synonyms are also listed in Table 2. For example: *shallow* can be replaced by its direct synonym *superficial*. The antonym of *shallow* is *deep*, and *deep* can be replaced by its direct synonym *profound*.

I extracted from the corpus all the sentences in which the negated and non-negated adjectives (which are listed in Table 2) appeared. An assistant and I independently examined each sentence and removed those in which the adjectives did not match the salient meanings upon which I had decided (and which are listed in Table 2). For instance, in the filtering process of *hard*, “... arguably brought on by *hard* liquor”, “*hard* rocking” and “... he is interested in learning the *hard* facts about ...” were removed. Then the filtered results were compared. For a sentence to be included in the count, both the assistant and I had to totally agree on its compatibility with the definition that appears in Table 2; 96% of the sentences selected in the first filtering stage survived the second filtering stage.

---

<sup>10</sup> Technically speaking, the globally salient meaning of an item in WordNet is the first definition of the item.

**Table 3:** SI values of the 16 adjectives forming the 8 canonical adjectival pairs

| Pair | Adjective | SI    | Adjective   | SI    |
|------|-----------|-------|-------------|-------|
| 1    | Bad       | 0.321 | Good        | 0.056 |
| 2    | Shallow   | 0.240 | Deep        | 0.007 |
| 3    | Wrong     | 0.198 | Right       | 0.080 |
| 4    | Ugly      | 0.183 | Beautiful   | 0.001 |
| 5    | Boring    | 0.109 | Interesting | 0.025 |
| 6    | Sad       | 0.090 | Glad        | 0.003 |
| 7    | Hard      | 0.071 | Easy        | 0.052 |
| 8    | Expensive | 0.032 | Cheap       | 0.004 |

As previously explained, for each adjectival pair (and their direct synonyms) I considered the group of { unfavorable adjective, negated favorable antonym } and the group of { favorable adjective, negated unfavorable antonym }. Pairs appearing less than 50 times were removed. Thus, although making an excellent canonical pair (identified as such by Deese 1964; Paradis 2010; and Van de Weijer et al. 2012), the *clean-dirty* pair was removed from the count, since even though { *clean, not dirty* } were counted altogether 337 times, { *dirty, not clean* } were counted 18 times only.

I then calculated the SI value for each adjective. The SI results appear in Table 3. Note that SI values calculated for unfavorable adjectives (ranging between 0.321 and 0.032) are considerably larger than those calculated for favorable adjectives (ranging between 0.080 and 0.001), ( $W(7) = 36$ ,  $p < 0.05$ ), meaning that unfavorable adjectives are more extensively replaced by their negated favorable antonyms (e.g., *bad* replaced by *not good*) than favorable adjectives replaced by their negated unfavorable antonyms (e.g., *good* replaced by *not bad*), and suggesting that a negated favorable antonym could potentially mitigate a unfavorable adjective, whereas a negated unfavorable antonym may not necessarily mitigate a favorable adjective.

## 6 Behavioral data

SI values should be compared against the gold standard, that is, human ratings of the polarity strength of all the adjectives for which the SI was calculated. Being aware of the extensive work using semantic differential scales (Osgood et al. 1957), I searched for listings of concepts which were rated for polarity strength, such as in Jenkins (1960) and in Jenkins et al. (1958), but I came up with partial

lists in which only a few of the 16 adjectives of interest existed. I, therefore, compiled a list of the Hebrew equivalents<sup>11,12</sup> of the target adjectives together with filler adjectives. Participants were presented with the list of adjectives and asked to rate each adjective on a -5 to 5 evaluative scale according to its strength.

## 6.1 Method

### 6.1.1 Participants

One-hundred and seventy undergraduate students of *AFEKA* – Tel-Aviv Academic College of Engineering (24 women and 146 men), mean age 26.18 (SD = 3.79), volunteered to participate in the experiment. They were all native speakers of Hebrew and had normal or corrected-to-normal vision.

### 6.1.2 Material

There were 32 adjectives translated to Hebrew, of which 16 were the target adjectives, used for the SI calculations, and 16 additional adjectives were filler items,

---

**11** It is preferable, no doubt, to compare SI values extracted from a corpus written in a certain language with speakers' evaluation of the potency of adjectives in the very same language. This was, however, impossible due to the absence of such an evaluative corpus in Hebrew. Regardless, the current cross-linguistic comparison relies on Osgood et al. (1975) as well as on Heise (2000), who provided strong evidence that people across cultures and languages rate similar concepts on the potency scale in the same way. Somewhat related major cross-cultural psychological studies, in which native Hebrew speakers participated (among other language speakers), suggest that personality trait structure (McCrae and Costa 1997) and the meaning interpretation of particular values (Schwartz 1992) may be universal.

**12** The common practice in translating questionnaires from one language to another language is Brislin's (1970) method, in which the translation from the source language to the target language and the back-translation to the source language are compared, trying to preserve the meaning of the items while performing minor literal changes. Then, the translated questionnaire is pretested on the targeted population and the results are compared with the results of the source population. Following Brislin, I translated the 16 adjectives of interest from Hebrew to English corroborating my translation by the most comprehensive English-Hebrew dictionary (Alcalay 1996). Then a Hebrew-to-English translator was asked to translate the adjectives back into English. We achieved full agreement, which is not surprising, given the canonical nature of the adjectives. No pretest on the targeted population and comparison to the results of the source population could be performed due to the small number of native English speakers willing to participate in the rating experiment.



## 6.2 Results and discussion

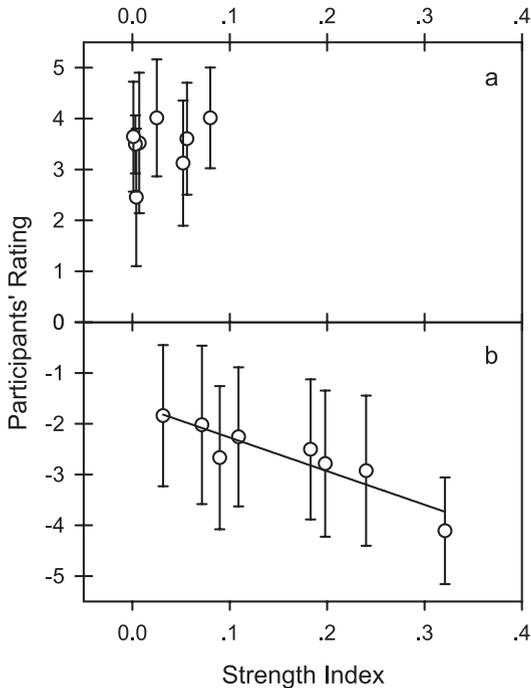
Ratings of one participant were discarded from the analysis because he acted automatically, rating all adjectives as 5, regardless of their polarity. The results of the ratings are presented in Table 4. They show that, as expected, favorable adjectives were rated along the positive portion of the rating scale, while unfavorable adjectives were rated along the negative portion of the scale. In addition, as expected, the standard deviations of the mean ratings of most adjectives are fairly large. These large standard deviations stem from the fact that the items were presented in isolation, such that each participant could activate a different context: It stands to reason to assume that an *ugly* face would evoke a different interpretation than an *ugly* purse or an *ugly* car, and thus participants' ratings of the polarity strength of *ugly* are likely to vary across potential contexts.

**Table 4:** Participants' ratings of the 16 adjectives forming the 8 canonical adjectival pairs

| Pair | Adjective | Participants' rating<br>M (SD) | Adjective   | Participants' rating<br>M (SD) |
|------|-----------|--------------------------------|-------------|--------------------------------|
| 1    | Bad       | -4.11 (1.05)                   | Good        | 3.60 (1.10)                    |
| 2    | Shallow   | -2.92 (1.48)                   | Deep        | 3.49 (1.38)                    |
| 3    | Wrong     | -2.78 (1.44)                   | Right       | 4.01 (0.99)                    |
| 4    | Ugly      | -2.55 (1.38)                   | Beautiful   | 3.64 (1.08)                    |
| 5    | Boring    | -2.26 (1.37)                   | Interesting | 4.01 (1.15)                    |
| 6    | Sad       | -2.67 (1.41)                   | Glad        | 4.58 (0.58)                    |
| 7    | Hard      | -2.02 (1.56)                   | Easy        | 3.12 (1.24)                    |
| 8    | Expensive | -1.84 (1.39)                   | Cheap       | 2.45 (1.35)                    |

Figure 2 presents the participants' ratings vs. the SI results calculated for the 16 adjectives of interest. Spearman's correlation between participants' ratings of unfavorable adjectives and the SI results is statistically significant,  $r_s(1350) = -0.39$  ( $t(1350) = -15.56$ ,  $p < 0.001$ ). However, Spearman's correlation between participants' ratings of favorable adjectives and the SI results is statistically not significant,  $r_s(1350) = 0.03$  ( $t(1350) = 1.04$ ,  $p = 0.30$ ).

The moderately strong correlation of speakers' evaluation of the strength of unfavorable adjectives with their SIs, but the lack of correlation of speakers' evaluation of the strength of favorable adjectives with their SIs match pattern 1 (see above). This correlation pattern shows that *bad* and *not good* delineate a single sample space. Thus, it stands to reason that a negated favorable antonym (e.g., *not good*) serves as the mitigated version of an unfavorable adjective (e.g., *bad*);



**Fig. 2:** Participants' ratings of the strength of the 16 adjectives vs. the adjectives' calculated SI values: (a) Spearman's correlation between participants' ratings of favorable adjectives and the SI results is statistically not significant,  $r_s(1350) = 0.03$  ( $t(1350) = 1.04$ ,  $p = 0.30$ ). (b) Spearman's correlation between participants' ratings of unfavorable adjectives and the SI results is statistically significant,  $r_s(1350) = -0.39$  ( $t(1350) = -15.56$ ,  $p < 0.001$ )

*good* and *not bad*, however, do not delineate a single sample space. Thus, a negated unfavorable antonym (e.g., *not bad*) is not considered as the mitigated version of a favorable adjective (e.g., *good*). Pattern-1-type correlation found here indicates that the concept in the scope of negation is indeed retained, given supporting positive expectations.

## 7 Summary and conclusions

This article aimed at providing a quantitative corpus-based test for the non-obligatoriness of suppression of negated concepts, also referred to as the negation as mitigation hypothesis (Giora 2003; Giora et al. 2005b), via a two-pronged methods which combines corpus data with behavioral data.

For the purpose of this study, I *intentionally naively* assumed that a canonical polar adjective and its negated antonym are logically equivalent; hence a sample space from which either of them could be chosen was delineated. I further assumed that using the negated antonym instead of the explicit adjective depends (at least, partially) on the strength of the particular adjective. The stronger an adjective, the more infrequent or shunned it is, hence replaceable by a mitigated (i.e., negated) alternative. A statistic – the Strength Index (SI) – which allows quantifying these assumptions was formulated: The denominator of the SI – which is the total number of occurrences of the adjective as well as its negated antonym within the corpus – represents the availability of both the adjective and its negated antonym in memory. The numerator of the SI (of an adjective) – which is the number of times the negated antonym (of an adjective) occurs within the corpus – represents the availability of the negated option.

SI values were calculated (Table 3) for 16 polar adjectives, forming 8 canonical antonymous pairs (such as *good-bad*, *right-wrong*, *cheap-expensive*, etc.). For the sake of reliable statistics, the salient meanings of the adjectives were referred to.

Then, an experiment was run in which participants rated the 16 adjectives (presented in isolation) on a –5 to 5 evaluative scale. This behavioral data and the calculated SI values were correlated (for unfavorable adjectives such as *bad*, and then for favorable adjectives such as *good*). Given speakers' prior positive expectations, correlations were explained in terms of retention or suppression of the concept in the scope of negation:

The adjectives and their negated antonyms, which I used for calculating the SI, were extracted out of a corpus which is made up of texts written by authors who had prior positive expectations. In light of these positive prior expectations, retention of the concept in the scope of negation is supported by correlation between the speakers' evaluation of the strength unfavorable adjectives and their SI values, but the lack of correlation between speakers' evaluation of the strength of favorable adjectives and their SI values. Specifically, favoring a negated favorable description (e.g., *not good*) over an alternative unfavorable description (e.g., *bad*), when a negative evaluation is in order, is allowed by retaining in memory the representation of the favorable adjective (e.g., *good*) appearing in the scope of negation; similarly, when a positive evaluation is in order, the negated unfavorable description (e.g., *not bad*) will not be favored over a favorable description (e.g., *good*) and will not serve as its substitute, again, only because the unfavorable adjective (e.g., *bad*) within the scope of negation is retained in memory. In both cases the SI captures the relation between an adjective and its negated antonym as a result of the retention of the latter in the scope of negation.

To the best of my knowledge, this study is the first one that tackles the issue of retention or suppression of a concept in the scope of negation via corpus. As

such it replicates Colston's (1999) results, which show that when assuming positive expectations, *not good* is preferred over *bad* while *not bad* cannot replace *good*. This asymmetry has been also found earlier in a somewhat different context. In an effort to design reliable semantic differential rating scales, Gilpin (1973) wanted to provide evidence that the asymmetry of semantic differential scales derives from lexical marking. He asked participants to rate concepts on three types of rating scales: bipolar scales such as *good-bad*; unipolar scales defined by the favorable member of a pair of antonyms, such as *good-not good*; and unipolar scales defined by the unfavorable member of a pair of antonyms, such as *not bad-bad*. Gilpin found that ratings of concepts on the bipolar scale bore stronger resemblance to the ratings on the unipolar scale, defined by the favorable member of the pair of antonyms than to the ratings on the unipolar scale, defined by the unfavorable member of the pair of antonyms. That is, processing of scale ends such as *good-bad* resembles processing of scale ends such as *good-not good* to a greater extent than *not bad-bad*. In other words, participants interpreted *not good* as closer to *bad* than *not bad* to *good*.

Production experiments such as Mann's (1968) support this asymmetry too (as well as the strong relation between production and comprehension processes referred to in Section 1). Mann hypothesized that, compared to favorable adjectives, unfavorable adjectives encourage definitions that deny the opposite. Accordingly, he asked one group of participants to define a selection of favorable adjectives and another group of participants to define a selection of unfavorable ones. Mann demonstrated that definitions by denial were not abundant for either unfavorable (e.g., *bad*) or favorable adjectives (e.g., *good*), since "to define *courageous* as *not cowardly* would be to leave open the possibility that innumerate qualities other than cowardliness are being referred to" (p. 766). Yet, definitions by denial were more readily provoked by unfavorable adjectives than by favorable adjectives. This observation could be accounted for by the retention of the concept in the scope of negation.

For the research done here to be complete, it would be necessary to conduct an additional experiment in which corpora written under prior negative expectations would be used. A correlation pattern which is a mirror image of pattern 1 is expected, i.e., correlation between speakers' evaluation of the strength favorable adjectives (e.g., *good*) and SI values, but lack of correlation between speakers' evaluation of the strength of unfavorable adjectives (e.g., *bad*) and SI values. To the best of my knowledge, no such corpora are currently available.

All in all, the preference of *not good* over *bad*, and *good* over *not bad*, indicates that the concept in the scope of negation is not obligatorily suppressed. Instead, it is retained as a result of contextually-dependent expectations.

## Appendix

**Table 5:** The 32 adjectives presented to the participants, their Hebrew translations in standard IPA, and the additional information provided to the participants in order to avoid misinterpretation of items. The first 16 adjectives are the adjectives used in the SI calculations, while the rest are filler items

| #  | Adjective        | Hebrew translation<br>in standard IPA<br>(Common alternatives:<br>f = š, ts = c, j = y, ʁ = r) | Additional information<br>provided to the participants |
|----|------------------|--|--|
| 1  | Right            | tsodek   |  |
| 2  | Cheap            | zol  | (Price-wise)   |
| 3  | Good             | tov  |  |
| 4  | Happy-Glad       | sameʔax  |  |
| 5  | Beautiful-Pretty | jafe   | (Physically)   |
| 6  | Easy             | kal  | (As in simple)   |
| 7  | Interesting      | meʔanjen   |  |
| 8  | Deep             | ʔamok  | (In the abstract sense)                                |
| 9  | Bad              | ʁa   |  |
| 10 | Expensive        | jakaʁ  | (Price-wise)   |
| 11 | Wrong            | ʃaguj  |  |
| 12 | Hard             | kafe   | (As in complicated)                                    |
| 13 | Ugly             | mexoʔaʁ  | (Physically)   |
| 14 | Sad              | ʔatsuv   |  |
| 15 | Shallow          | fitxi  | (In the abstract sense)                                |
| 16 | Boring           | meʃaʔamem  |  |
| 17 | Accurate         | medujak  |  |
| 18 | Coward           | paxdan   |  |
| 19 | Creative         | jetsiʁati  |  |
| 20 | Cruel            | ʔaxzaʁ   |  |
| 21 | Handy            | ʃimufi   |  |
| 22 | Dangerous        | mesukan  |  |
| 23 | Violent          | ʔalim  |  |
| 24 | Obedient         | tsajtan  |  |
| 25 | Studious         | ʃakdan   |  |
| 26 | Stupid           | tipeʃ  |  |
| 27 | Friendly         | jediduti   |  |
| 28 | Vague            | meʔuʁpal   | (In the abstract sense)                                |
| 29 | Innovative       | xadʃani  |  |
| 30 | Repulsive        | doxe   |  |
| 31 | Vivacious        | nimʁats  |  |
| 32 | Lonely           | boded  |  |

**Acknowledgments:** I am grateful to Stefan Th. Gries and two anonymous reviewers for their constructive comments on an earlier version of this article. I am indebted to my brother, Yonatan Becker, who served as my devoted assistant in meticulously filtering the adjectives' lists, as well as in handling many of the technicalities of this research. I am thankful to Aviad Albert for his useful suggestions regarding the transcriptions of the Hebrew items into standard IPA. All errors and oversights are, of course, my own. This research was supported by the Tel-Aviv University dean of humanities' scholarship for outstanding Master's students in linguistics.

## References

- Alcalay, Reuven. 1996. *The complete English-Hebrew dictionary*. Tel-Aviv: Chemed Books-Massada-Yedioth Ahronoth.
- Autry, Kevin S. & William H. Levine. 2012. Activation of negated and non-negated entities. *Journal of Pragmatics* 44. 1474–1485.
- Autry, Kevin S. & William H. Levine. 2014. Presupposition processing and the (re)activation of negated concepts. *Discourse Processes* 51(7). 535–564.
- Blitzer, John, Mark Dredze & Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. Paper presented at the Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 23–30 June.
- Bradac, James J., John W. Bowers & John A. Courtright. 1979. Three language variables in communication research: Intensity, immediacy, and diversity. *Human Communication Research* 5(3). 257–269.
- Brislin, Richard W. 1970. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology* 1(3). 185–216.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage* (studies in interactional sociolinguistics). New-York: Cambridge University Press.
- Colston, Herbert L. 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.
- de Jung, Ilona K. E. & Christian Burgers. 2013. Do consumer critics write differently from professional critics? A genre analysis of online film reviews. *Discourse, Context & Media* 2(2). 75–83.
- Deese, James. 1964. The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior* 3(5). 347–357.
- Deese, James. 1965. *The structure of associations in language and thought*. Baltimore: The Johns Hopkins Press.
- Esuli, Andrea & Fabrizio Sebastiani. 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining. Paper presented at the International Conference on Language Resources and Evaluation, Genoa, Italy, 22–28 May.
- Fraenkel, Tamar & Yaacov Schul. 2008. The meaning of negated adjectives. *Intercultural Pragmatics* 5(4). 517–540.
- Franzwa, Helen H. 1969. Psychological factors influencing use of “evaluative-dynamic” language. *Speech Monograph* 36. 103–109.

- Gilpin, Andrew R. 1973. Lexical marking effects in the semantic differential. *The Journal of Psychology* 85(2). 277–285.
- Giora, Rachel. 2003. *On our mind: Salience, context, and figurative language*. New York: Oxford University Press.
- Giora, Rachel. 2006. Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics* 38. 981–1014.
- Giora, Rachel, Noga Balaban, Ofer Fein & Inbar Alkabetz-Zlozover. 2005a. Negation as positivity in disguise. In Herbert L. Colston & Albert N. Katz (eds.), *Figurative language comprehension: Social and cultural influences*, 233–258. Hillsdale: Erlbaum.
- Giora, Rachel & Ofer Fein. 1999. On understanding familiar and less-familiar figurative language. *Journal of Pragmatics* 31. 1601–1618.
- Giora, Rachel, Ofer Fein, Keren Aschkenazi & Inbar Alkabetz-Zlozover. 2007. Negation in context: A functional approach to suppression. *Discourse Processes* 43(2). 153–172.
- Giora, Rachel, Ofer Fein, Jonathan Ganzi, Natalie Alkesslasy Levi & Hadas Sabah. 2005b. On negation as mitigation: The case of negative irony. *Discourse Processes* 39(1). 81–100.
- Godbole, Namrat, Manjunath Srinivasaiah & Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs. Paper presented at the International Conference on Weblogs and Social Media, Boulder, CO, 26–28 March.
- Gross, Derek, Ute Fischer & George A. Miller. 1989. The organization of adjectival meanings. *Journal of Memory and Language* 28. 92–106.
- Hamilton, Mark A., John E. Hunter & Michael Burgoon. 1990. An empirical test of an axiomatic model of the relationship between language intensity and persuasion. *Journal of Language and Social Psychology* 9(4). 235–255.
- Hasson, Uri & Sam Glucksberg. 2006. Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics* 38. 1015–1032.
- Heise, David R. 2000. Affect control theory across cultures. Paper presented at the Annual Meetings of the American Sociological Association, Washington, DC, 12–16 August.
- Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.
- Jenkins, James J. 1960. Degree of polarization and scores on the principal factors for concepts in the semantic atlas study. *The American Journal of Psychology* 73(2). 274–279.
- Jenkins, James J., Wallace A. Russell & George J. Suci. 1958. An atlas of semantic profiles for 360 words. *The American Journal of Psychology* 71(4). 688–699.
- Jensen, Matthew L., Joshua M. Averbek, Zhu Zhang & Kevin B. Wright. 2013. Credibility of anonymous online product reviews: A language expectancy perspective. *Journal of Management Information Systems* 30(1). 293–324.
- Jespersen, Otto. 1917. *Negation in English and other languages*. Copenhagen: A. F. Høst.
- Jones, Steven, Carita Paradis, Lynne M. Murphy & Caroline Willners. 2007. Googling for ‘opposites’: A web-based study of antonym canonicity. *Corpora* 2(2). 129–154.
- Kamps, Jaap, Maarten Marx, Robert J. Mokken & Maarten de Rijke. 2004. Using WordNet to measure semantic orientations of adjectives. Paper presented at the International Conference on Language Resources and Evaluation, Lisbon, Portugal, 24–30 May.
- Kaup, Barbara, Jana Lüdtkke & Rolf A. Zwaan. 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38. 1033–1050.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge: The MIT Press.

- Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.
- MacDonald, Maryellen C. & Marcel Adam Just. 1989. Changes in activation level with negation. *Journal of Experimental Psychology* 15(4). 633–642.
- Mann, William C. 1968. Defining the unfavorable by denial. *Journal of Verbal Learning and Verbal Behavior* 7. 760–766.
- Mayo, Ruth, Yaacov Schul & Eugene Burnstein. 2004. “I am not guilty” vs “I am innocent”: Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology* 40. 433–449.
- McCrae, Robert R. & Paul T. Costa Jr. 1997. Personality trait structure as a human universal. *American Psychologist* 52(5). 509–516.
- Miller, George A. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography* 3(4). 235–244.
- Osgood, Charles E., William H. May & Murray S. Miron. 1975. *Cross-cultural universals in affective meaning*. Urbana: University of Illinois Press.
- Osgood, Charles E., George J. Suci & Percy Tannenbaum. 1957. *The measurement of meaning*. Urbana: University of Illinois Press.
- Paradis, Carita. 2010. Good, better and superb antonyms: A conceptual construal approach. *The Annual of Texts by Foreign Guest Professors*. Prague: Charles University in Prague, Faculty of Philosophy and Arts.
- Paradis, Carita & Caroline Willners. 2006. Antonymy and negation – The boundedness hypothesis. *Journal of Pragmatics* 38. 1051–1080.
- Pickering, Martin J. & Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36. 329–392.
- Pornpitakpan, Chanthika 2004. The persuasiveness of source credibility: A critical review of five decades’ evidence. *Journal of Applied Social Psychology* 34(2). 243–281.
- Schwartz, Shalom H. 1992. Universals in the content and structure of values: Theory and empirical tests in 20 countries. In Mark Zanna (ed.), *Advances in experimental social psychology*, vol. 25, 1–65. New York: Academic Press.
- Shuval, Noa & Barbara Hemforth. 2008. Accessibility of negated constituents in reading and listening. *Intercultural pragmatics* 5(4). 445–469.
- Slater, Michael D. & Donna Rouner. 1996. How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly* 73(4). 974–999.
- Sperber, Dan & Deirdre Wilson. 1986/1995. *Relevance: Communication and cognition*. Oxford: Basil Blackwell.
- Stephens, Greg J., Lauren J. Silbert & Uri Hasson. 2010. Speaker–listener neural coupling underlies successful communication. *PNAS* 107(32). 14425–14430.
- Stone, Philip J., Dexter C. Dunphy, Marshal S. Smith & Daniel M. Ogilvie. 1966. *The General Inquirer: A computer approach to content analysis*. Cambridge: The MIT Press.
- Taboada, Maitte. 2011. Stages in an online review genre. *Text & Talk* 31(2). 247–269.
- Tottie, Gunnel. 1991. *Negation in English speech and writing: A study in variation* (Quantitative Analysis of English Structure). San-Diego: Academic Press.
- Tottie, Gunnel & Carita Paradis. 1982. From function to structure: Some pragmatic determinants in impromptu speech. Paper presented at the Symposium on Problems in the Linguistic Study of Impromptu Speech, Abo, Finland, 20–22 November.
- Van de Weijer, Joost, Carita Paradis, Caroline Willners & Magnus Lindgren. 2012. As lexical as it gets: The role of co-occurrence of antonyms in a visual lexical decision experiment. In

Dagmar Divjak & Stefan Th. Gries (eds.), *Frequency effects in cognitive linguistics*, 255–279. Berlin: Mouton de Gruyter.

Williams, Gbolahan K. & Sarabjot S. Anand. 2009. Predicting the polarity strength of adjectives using WordNet. Paper presented at the International Conference on Weblogs and Social Media, San Jose, CA, 17–20 May.

## Bionote

*Israella Becker* is a Master's student at the Department of Linguistics, Tel-Aviv University. She is mostly interested in the psycholinguistics and pragmatics of negation. Apart from a Major in linguistics, she holds a PhD in chemistry. Both are from Tel-Aviv University.

Copyright of Corpus Linguistics & Linguistic Theory is the property of De Gruyter and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.