Suppression of Concepts in the Scope of Negation is *not* Obligatory –
A Corpus-Based Study

Israela Becker

According to the *Suppression Hypothesis*, negation is an instruction from a speaker to an addressee to reduce the activation levels of a concept to baseline or below baseline, and replace it with an alternative (e.g., Hasson and Glucksberg 2006; Kaup et al. 2006; MacDonald and Just 1989; Mayo et al. 2004). A different approach – the *Retention Hypothesis*, introduced by Giora (Giora 2006; Giora et al. 2005a, 2005b; Giora et al. 2007) – adopts a functional, context-based view. Accordingly, the negation operator does not obligatorily *suppress* the concept in its scope; instead it either *retains* it or suppresses it pending contextual considerations.

Activation levels of concepts, which indicate their *retention* or *suppression* in memory, are often monitored via on-line experiments: Subjects are presented a negative stimulus, following which, they have to make a lexical decision as to whether a probe-word is a word or a nonword. Response times to the probes indicate the activation levels of the probed concept.

Instead of using on-line experiments, the current research aims at providing a two-pronged method, combining corpus-based data and off-line experimental data in an attempt to test the aforementioned hypotheses. The method introduced here is based on analyzing the relation between an adjective of an emotive nature (e.g., *bad*) and its negated antonym (e.g., *not good*) – whether the negated antonym is a mitigated version of the adjective, or not. The specific tool introduced in order to examine this relation—and consequently test the aforementioned hypotheses—is the *Strength Index* (*SI*) which 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:

\[
SI_{Adjective} = \frac{(Negated \_Antonym)}{(Negated \_Antonym + Adjective)}
\]

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 its negated antonym (e.g., *not good*) appears in the same corpus. The ratio between the numerator and the denominator expresses the extent to which the Negated Antonym is preferred over the Adjective. The higher the SI, the less preferred (and hence, more replaced) the adjective (e.g., *bad*) with respect to its negated antonym (e.g., *not good*). In other words, the higher the SI, the stronger the adjective.

The validity of the SI as a measure of strength of an adjective draws on the relation between the adjective (e.g., *bad*) and its negated antonym (e.g., *not good*); the SI expresses the strength of an adjective only if its negated antonym serves as the mitigated version of the adjective. Specifically, correlation between SI values and speakers’ evaluation of strengths of adjectives indicates that a 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 SI values and speakers’ evaluation of strengths of adjectives 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 its lack thereof 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 depend on prior discourse expectations.

SI values were calculated and participants’ ratings of the strengths of adjectives were collected for favorable adjectives (e.g., *good*) and then for unfavorable adjectives (e.g., *bad*) of an emotive nature. The four possible correlation patterns presented in Table 1 will be discussed in light of prior discourse expectations, and evidence supporting Giora’s Retention Hypothesis will be presented.
Table 1: Possible correlation patterns between SI values and participants’ rating of strengths of adjectives of an emotive nature.

<table>
<thead>
<tr>
<th></th>
<th>Correlation between SIs of favorable adjectives and participants' rating</th>
<th>Correlation between SIs of unfavorable adjectives and participants' rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Suppression</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Suppression-Retention</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Retention-Suppression</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

References


