



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Pragmatics

journal homepage: [www.elsevier.com/locate/pragma](http://www.elsevier.com/locate/pragma)

## Absolutely relative or relatively absolute? The linguistic behavior of gradable adjectives and degree modifiers

Naomi Kamoen<sup>a,\*</sup>, Bregje Holleman<sup>a</sup>, Rick Nouwen<sup>a</sup>, Ted Sanders<sup>a</sup>, Huub van den Bergh<sup>a,b</sup>

<sup>a</sup> Utrecht Institute of Linguistics OTS, Utrecht University, 3512 JK Utrecht, Netherlands

<sup>b</sup> Graduate School of Teaching and Learning, University of Amsterdam, Netherlands

## ARTICLE INFO

## Article history:

Received 1 March 2010

Received in revised form 21 May 2011

Accepted 31 May 2011

Available online 12 July 2011

## Keywords:

Gradable adjectives

Degree modifiers

Surveys

Wording effects

Antonyms

## ABSTRACT

Respondents are more likely to disagree with negative survey questions (*This text is boring. Yes/No*) than to agree with positive ones (*This text is interesting. Yes/No*). The size of this effect, however, varies largely between word pairs. A semantic classification of adjectives in closed scale/absolute and open scale/relative types was predicted to explain this variation. To classify survey adjectives, a judgment experiment was conducted. Language users ( $N = 173$ ) rated sentences in which an adjective was modified by the maximizer *completely* or the approximator *almost*: it should be possible to combine closed scale/absolute adjectives with these modifiers, in contrast to open scale/relative adjectives for which this is not the case.

Results show that language users agree on which adjective and degree modifier combinations are acceptable and which combinations are unacceptable. Moreover, the two methods, *almost* and *fully*, show convergent validity. However, the rating of the same combination of a specific adjective and a specific degree modifier varies across contexts. This suggests that neither of the two methods allows for an unambiguous classification of adjectives. Hence, the distinction between closed scale/absolute and open scale/relative adjectives cannot explain variation in survey response effects. For semantics and pragmatics results indicate that context plays a crucial role in the linguistic behavior of adjectives and degree modifiers.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

The choice for a positive or a negative wording matters in all sorts of language use situations. For example: when ground beef is described as 75% *lean* it tastes better than when being described as 25% *fat* (Levin and Gaeth, 1988). Women are more prone to engage in breast self-examination when confronted with information stressing the negative effects of abstaining from breast self-examination, than when presented with information stressing the positive consequences of engaging in breast self-examination (Meyerowitz and Chaiken, 1987). And in surveys, respondents are more likely to answer *no* when asked whether speeches against democracy should be *forbidden*, than to answer *yes* when asked whether speeches against democracy should be *allowed* (e.g. Rugg, 1941).

The above are all examples of so-called framing effects (cf. Levin et al., 1998). In this article, we will focus on framing effects in attitude surveys. These have been investigated extensively since they were first discovered by Rugg in the 1940s (see the example above; cf. Bishop et al., 1988; Hippler and Schwarz, 1986; Holleman, 2000; Kamoen et al., 2007;

\* Corresponding author. Tel.: +31 30 253 8099.  
E-mail address: [N.Kamoen@uu.nl](mailto:N.Kamoen@uu.nl) (N. Kamoen).

Krosnick and Schuman, 1988; Schuman and Presser, 1981; Waterplas et al., 1988). The main conclusion to be drawn from these studies is that, as in Rugg's study, respondents are generally more likely to disagree with negative survey questions than to agree with positive ones. In other words: respondents express their opinions more positively when the question is worded negatively.

In spite of a general tendency for negative questions to be answered more positively, survey studies also show there is a large variation in the size of this effect: it is not the case that a wording effect in the expected direction is observed for every manipulated question in every experiment. This variation is partially attributable to differences in the question content and factors of the communicative context that are varied between studies (e.g. Bishop et al., 1988; Holleman, 1999). In addition, Kamoen et al. (submitted for publication) show that when generalizing over study and question characteristics, variation between word pairs exists too. In this latter study, wording effects for thirteen contrastive word pairs were investigated. Each contrastive word pair was examined in a range of about fifteen experimental settings. Results show that for some pairs an effect of question wording can be generalized across experimental settings: respondents express their opinions more positively when the question is worded negatively (e.g. *simple/complicated*; *appealing/distant*). For other word pairs the wording effect was not consistent across experimental settings (e.g. *clear/unclear*; *logical/illogical*).

An important issue raised by these results is: how can between word pair variation in response effects for contrastive attitude questions be explained? Several survey researchers have proposed that semantic characteristics of words may explain part of this variation (e.g. Hayes, 1939; Molenaar, 1982; Holleman, 2000). In the current study, a first step will be taken to apply the study of semantics to the field of survey research: it will be investigated whether or not adjectives often used in surveys, such as *easy/difficult* and *clear/unclear*, can be classified unambiguously into different semantic types.

The semantics of contrasts have been studied since Aristotle. During the past decade, the study of gradable adjectives has gained renewed interest (Croft and Cruse, 2004; Kennedy, 1999, 2001, 2007; Kennedy and McNally, 2005; Paradis, 1997, 2000, 2001, 2008; Paradis and Willners, 2006; Rotstein and Winter, 2004; Syrett, 2007; Tribushinina, 2008, 2009). Working from different linguistic perspectives, several researchers have proposed similar ways in which adjectives can be classified into different types. In the current study, we will adopt the degree perspective as starting point (e.g. Kennedy and McNally, 2005; Kennedy, 2007; Rotstein and Winter, 2004),<sup>1</sup> to explain how adjectives can be distinguished into two semantic types based on the scale structure and the type of comparison standard.

On the one hand, there are antonyms such as *tall* and *short* that map their arguments onto an open scale lacking endpoints (e.g. Kennedy and McNally, 2005; Kennedy, 2007). Open scale adjectives are interpreted against a contextually determined "relative" comparison standard (e.g. Kennedy and McNally, 2005; Kennedy, 2007). For *tall* and *short*, this is illustrated by the fact that both an adult and a toddler can be referred to as *tall*, even though a *tall adult* is much taller than a *tall toddler* is.

On the other hand, there are adjectives that map their arguments onto (partially) closed scales that do have endpoints on either or both sides (e.g. Kennedy and McNally, 2005; Kennedy, 2007). For the pair *bent/straight*, *straight* is associated with a maximum element that is ordered above all others. Intuitively this follows from the observation that, for example, a line can be *maximally straight*. Hence, the scale of *bent/straight* is partially closed. Fully closed scales, of which the scale for *full/empty* is an example, are less common. Such scales are associated with two maximal endpoints

Adjectives that map their arguments onto (partially) closed scales are interpreted against a context independent standard of comparison. There are two subtypes of such absolute comparison standards. Maximum standard adjectives, such as *straight*, *full*, and *empty* always describe a maximal degree of the relevant property or, put differently, a zero degree of the opposite property. For example, if a stick is *straight*, this roughly means that the stick is 'free of bend' (compare Kennedy, 2007). This rough definition applies to the adjective irrespective of what exactly is being described as *straight*, for example, a *stick* or a *human's back*. Minimum standard adjectives such as *bent* always describe 'a non-zero degree of a relevant property': if a stick is *bent* this roughly means that it has 'non-zero bend' (compare Kennedy, 2007).<sup>2</sup>

In the search for an explanation of between word pair variation in framing effects, the fact that open scale/relative and closed scale/absolute word pairs are associated with different entailment patterns (e.g. Cruse, 1986; Kennedy and McNally, 2005; Rotstein and Winter, 2004), is particularly interesting. For open scale/relative adjectives, there is no one-on-one relation between the denial of the one pole and the affirmation of the other. However, for closed scale/absolute word pairs, this one-on-one relation does exist. This is shown in (1; example taken from Kennedy and McNally, 2005:359) and (2).

- (1) The door is not large (small)  $\not\equiv$  The door is small (large)
- (2) The stick is not bent (straight)  $\equiv$  The stick is straight (bent)

Applied to survey contexts, this means it is to be expected that a larger difference in answers occurs for contrastive questions with relative adjectives. As for relative adjectives there is no one-on-one relation between the denial of the one pole and the affirmation of the other, this implies that someone who answers "no" to a positive question with a relative adjective, does not

<sup>1</sup> It is a relatively dominating view in semantics that "gradable adjectives map their arguments onto abstract representations of measurement or DEGREES. (...) A set of degrees totally ordered with respect to some DIMENSION (height, cost, etc.) constitutes a SCALE" (Kennedy, 2007:4). This means that, for example, in the sentence *John is tall*, the argument John is mapped onto some degree on the scale of the dimension of height.

<sup>2</sup> These examples seem to imply that only maximum/maximum standard word pairs and minimum/maximum standard word pairs exist. However, also pairs of two minimum standard adjectives exist. Such pairs are not discussed here.

necessarily have to answer “yes” to the same question worded negatively. For absolute adjectives this one-on-one relation does exist, which suggests that respondents may be more likely to give the same answer to positive and negative questions.

To test whether this typology of adjectives explains variation in survey response effects, sets of open scale/relative and closed scale/absolute survey adjectives are required. Such sets, however, cannot be distilled from the semantic literature directly: survey adjectives such as *acceptable/unacceptable*, *easy/difficult* and *good/bad* are not often analyzed as exemplar cases of closed scale/absolute or open scale/relative adjectives. This means a method is required to distinguish between different types of survey adjectives.<sup>3</sup>

In the current study, two methods will be used for classifying survey adjectives. First, the linguistic behavior of adjectives and maximizers will be used as an empirical probe to determine the scale structure and the type of comparison standard (e.g. Kennedy and McNally, 2005; Kennedy, 2007). Maximizers such as *100%*, *fully*, and *completely* are degree modifiers that indicate that an endpoint has been reached. Maximum standard adjectives combine with maximizers. Therefore, as a linguistic repercussion of the scale structure, either or both members of a word pair associated with a closed scale should felicitously combine with this type of degree modifier. In contrast, neither of the two members of an open scale/relative word pair should combine with maximizers. This has been shown in (3) and (4).

- (3) Open scale structure, relative comparison standard:  
James is fully #tall/#short
- (4) Closed scale structure, absolute comparison standard:  
The stick is 100% #bent/straight

Second, the linguistic behavior of adjectives and approximators will also be used as a method to determine the scale structure and hence the type of comparison standard (Rotstein and Winter, 2004). Approximators are degree modifiers such as *almost* that indicate the approximation to an endpoint. As such, these modifiers combine with maximum standard adjectives. Hence, at least one of the members of a closed scale/absolute word pair should combine with these modifiers. By contrast, neither of the members of an open scale/relative word pair should do so (see examples 5 and 6).

- (5) Open scale structure, relative comparison standard:  
James is almost #tall/#short
- (6) Closed scale structure, absolute comparison standard:  
The stick is almost #bent/straight

In sum, in the current study, we will use two methods, that is, maximizers and approximators, to determine the type of scale structure and the type of comparison standard for a group of survey adjectives. Closed scale/absolute word pairs are those pairs for which at least one of the members can co-occur with maximizers and approximators. In contrast, open scale/relative adjectives are those adjectives for which neither of the members combines with maximizers and approximators. Please note that the use of two empirical probes is important to assess the convergent validity. From a semantic perspective, the use of two modifiers instead of one allows for an exploration of the extent to which the same adjectives do and do not combine with maximizers and approximators.

Crucial for the explanation of between word pair variation in survey framing effects is that the scale structure and the comparison standard are lexical characteristics of words: if the scale structure and the type of comparison standard are always uniquely determined in a certain context, no between word pair variation can be explained. It has generally been assumed that an adjective is associated with only one type of scale structure (open/closed), and that the absolute or relative interpretation of an adjective is by default attached to the scale structure (e.g. Kennedy and McNally, 2005; Kennedy, 2007). The role of such conventional elements should always be maximized (Kennedy, 2007). Of course this implies there are non-default situations in which context overrules the standard interpretation of an adjective.<sup>4</sup> Nevertheless, this view necessarily also implies that for one and the same adjective the scale structure and type of comparison standard will be relatively stable

<sup>3</sup> Beforehand, there is no reason to assume that survey adjectives cannot be classified into the mentioned types. Yet, in this respect, we would like to make a side step to the work of Bierwisch (1989). Bierwisch distinguishes a class of evaluative adjectives as opposed to the class of “ordinary” dimensional adjectives. Roughly speaking, dimensional adjectives measure some physical property of an object (e.g. *height*), whereas evaluative adjectives measure a subjective property (e.g. *beauty*). In semantic literature on the absolute/relative distinction, mainly dimensional adjectives are examined. However, some or most survey adjectives are probably of the evaluative type. Although evaluative adjectives may have distinguishing features (see Bierwisch, 1989), there is no reason to assume there is an a priori clash between this group of adjectives and the distinction between absolute/relative adjectives: the relation between these two classifications is usually not discussed explicitly in the literature. An open scale/relative comparison standard may by definition seem more obvious for adjectives that have an evaluative nature or are evaluative in use, however a set of survey adjectives (e.g. *clear/unclear*; *truthful/untruthful*) is analyzed as a closed scale absolute word pair by Rotstein and Winter (2004).

<sup>4</sup> There is no clear “rule” for how often or in exactly what situations context overrules the default interpretation of an adjective. With respect to the relation between the scale structure and the associated type of comparison standard, Kennedy and McNally (2005) note that open scale adjectives are always associated with a relative comparison standard. However, the relation between (partially) closed scales and an absolute comparison standard is not always one to one: the scalar endpoints of (partially) closed scales are in some contexts not used as standard of comparison (Kennedy and McNally, 2005: 361).

across a set of random contexts. For survey research it is important to investigate whether the scale structure and the type of comparison standard can be determined relatively unambiguously. Therefore, the linguistic behavior of the same adjective and degree modifier combination will be investigated in several linguistic contexts. In doing so, the current research adds to semantic theory by providing an empirical exploration of exactly how large or how small the role of contextual factors is in the linguistic behavior of gradable adjectives and degree modifiers.

## 2. Method

### 2.1. Participants

A large scale judgment experiment was carried out in which 173 participants rated the acceptability of statements with a gradable adjective modified by a maximizer or an approximator. The linguistic material was Dutch<sup>5</sup> and all raters were native speakers of Dutch. Most of them were female (75%). This sex ratio reflects the sex ratio at the Faculty of Humanities at Utrecht University where the experiment was conducted.

As the rater's linguistic background may influence acceptability ratings, two groups of raters were distinguished, based on language expertise. An advantage of distinguishing two groups of raters is that the linguistic behavior of adjectives and degree modifiers can be examined twice, and that possible differences between the two groups of raters can be explored.

The first group of raters consisted of 60 language experts. Language experts were defined as people who are capable of looking at the Dutch language from a meta-perspective. MA-students and PhD-students in Linguistics, Dutch language and literature, and Language and communication studies are assumed to have obtained a meta-view on language through their studies. The second group of raters consisted of 113 non-experts. These were all students who study a non-language related subject at the Humanities Faculty, such as Contemporary history.

### 2.2. Procedure

The raters first answered questions about their gender, and educational background. Then, they rated the acceptability of 40 different statements in which a combination of an adjective and a degree modifier occurred. To allow for a distinction between more and less acceptable sentences, the raters could express their opinion on a five-point scale, ranging from 1 (unacceptable) to 5 (acceptable). The instruction for judging the sentences was based on Tribushinina (2009):

"Please indicate whether the following sentences are acceptable Dutch in your opinion. A sentence is acceptable if it sounds natural and you would not be surprised to encounter it in daily language. A sentence is unacceptable if it sounds strange or is incorrect in a grammatical sense. You can express your opinion on a five-point scale. Please choose "1" if you think the sentence is completely unacceptable. Pick "2" if you think the sentence is unacceptable. Choose "3" if you think the sentence is neither acceptable nor unacceptable. Pick "4" if you think the sentence is acceptable. Pick "5" if you think the sentence is completely acceptable. Please remember, you should only base your rating on whether you think the sentence is acceptable in a linguistic sense, not on whether you agree with the content of the sentence."

The raters in this study participated voluntarily, or as a course requirement. It took them roughly 15 min to judge the 40 sentences. Even though this was not a lot of time, we did not increase the number of sentences, because we know from experience that linguistic intuitions decrease as the number of items to rate increases.

### 2.3. Materials

The set of 40 items each rater evaluated was always one of the 20 randomly selected sets of statements, which were drawn from a list of 800 statements. This list of statements was comprised as follows. First, 20 Dutch antonymous word pairs, that is, 40 adjectives, were selected. Second, 10 survey statements with each of these adjectives were constructed. Third, two versions of each statement were created: one with the adjective modified by the approximator *bijna* (almost), and one with the adjective modified by the maximizer *volledig* (fully). In the following sections we will elaborate on the choice of adjectives, contexts, and degree modifiers made while comprising this material.

#### 2.3.1. The choice of adjectives

Twenty antonyms were selected from an antonym list.<sup>6</sup> All of the selected adjectives are frequently used in attitude surveys. Examples are *duidelijk/onduidelijk* (clear/unclear), and *makkelijk/moeilijk* (easy/difficult). For all Dutch adjectives and their translations, see Appendix A.

<sup>5</sup> Kennedy and McNally (2005) note that the elements of their typology are fundamental enough to be applicable to other languages as well. In addition, Vanden Wyngaerd (2001) shows that the closed scale/absolute versus open scale/relative typology is indeed largely applicable to Dutch.

<sup>6</sup> <http://www.mijnwoordenboek.nl/antoniwm.php>.

### 2.3.2. The definition of context

Each adjective occurred in 10 different contexts. These were all sentences that could be used in an attitude survey. All statements concerned one issue: Dutch smoking policies. Known contexts in which non-default linguistic behavior occurs were avoided.<sup>7</sup> The statements only varied with respect to seemingly unimportant characteristics, such as their exact content (e.g. 'In the Netherlands, it is easy/difficult to buy cigarettes when you are under sixteen' versus 'In the Netherlands, it is easy/difficult to buy marijuana when you are under sixteen'). Moreover, adjectives were occasionally used predicatively (e.g. 'In my view, it is acceptable/unacceptable that people smoke in ice bars') and attributively (e.g. 'Increasing taxes on cigarettes is an acceptable/unacceptable way to discourage smoking'). The same contexts were used for the positive and the negative members of the word pairs. As an example, the 10 contexts for the word pair *easy/difficult* are provided in Appendix B.

### 2.3.3. The choice of degree modifiers

Two versions of each sentence were created: one with the adjective modified by the maximizer *fully* and one with the adjective modified by the approximator *almost*. The specific approximator *almost* was chosen because it is explicitly put forward in the work of Rotstein and Winter (2004). For the maximizer, there was a broader range of possible alternatives to chose from. Some maximizers, such as *helemaal* (*completely*), nowadays also allow for a reading as a booster in Dutch (Hoeksema, 2008). A booster is a degree modifier such as *very* that has a reinforcing effect on the modified property (Paradis, 2008:321). As such, it also combines with relative adjectives, which is a potential threat to the validity of the experiment. According to our own linguistic intuitions, *volledig* (*fully*) has no such reading in Dutch. Therefore, this specific maximizer was selected. All in all, the sentences to be rated were sentences such as the following examples: 'In my view, it is almost acceptable that people smoke in ice bars' and 'In my view, it is fully acceptable that people smoke in ice bars'. The negative members of the word pairs were always used in the same constructions, hence: 'In my view, it is almost unacceptable that people smoke in ice bars' and 'In my view, it is fully unacceptable that people smoke in ice bars'.

### 2.4. Pretest

To make sure that the possible unacceptability of the test sentences can only be ascribed to the combination of the adjective and the degree modifier, a pretest was conducted in which 60 language experts rated the acceptability of sets of sentences without the adjective modified by *almost* or *fully*. None of the experts in the pretest also participated in the main study. All sentences used in the pretest, were judged to be acceptable (reliability .97). Therefore, the possible non-acceptability of sentences in the main experiment, in which the adjective is modified by *almost* or *fully*, can only be ascribed to the addition of the degree modifier.

### 2.5. Analysis

It is important to stress that the data obtained in this experiment have a clear hierarchical structure (Quené and Van den Bergh, 2004, 2008). In this hierarchical structure, three sources of variance can be distinguished. First, there is a clustering of ratings of a specific adjective: the one adjective is more acceptable with *almost* and *fully* than the other adjective is. This source of variance will be referred to as the *word variance*. Second, there is a clustering of ratings of one rater, because one rater is stricter than the other. This source of variance will be called the *person variance*. Third, there is a clustering of ratings that can be attributed to the specific combination of an adjective and the rater who rated it. If the same rater judges a specific combination of an adjective and a degree modifier differently in different contexts, a third source of variance will be created. This source of variance will be called the *contextual variance*.

To do justice to the hierarchical structure of the data, a multi-level model was used for analyzing the data (cf. Quené and Van den Bergh, 2004, 2008). In this model, eight mean rating scores were estimated: one for each combination of the degree modifier (*almost* versus *fully*), the polarity of the adjective (positive versus negative),<sup>8</sup> and the type of rater (expert versus non-expert). Moreover, the model estimates word variance, person variance, and contextual variance for each of the eight mean scores (see Appendix C for a further explanation and formalization).

Using this model, the reliability of the ratings will be assessed first. It is important that the reliability of the ratings is acceptable, because if not, there is in fact no linguistic behavior to analyze. Second, the convergent validity for the two methods will be determined. If there is convergent validity, a strong positive correlation between ratings of sentences with

<sup>7</sup> There are some known special circumstances under which non-default linguistic behavior occurs. For example, if the standard of comparison is made explicit in the context, relative terms do combine with approximators: *A tall basketball player is someone above 2.00 meters high. John is 1.98 meters, so he is almost tall* (example taken from Rotstein and Winter, 2004:276). To account for this problem, we avoided making the standard of comparison explicit in our materials. Also, *bijna* sometimes gets a temporal reading and therefore combines with some color terms (compare Amaral, 2006). To circumvent this problem, we avoided color terms in our study. We would like to stress that a temporal reading is not likely to occur for the kind of adjectives we used in the current study.

<sup>8</sup> With positive and negative we refer to "evaluatively positive" and "evaluatively negative" (Hamilton and Deese, 1971). Note that the definition of positive and negative is in fact irrelevant here; what is important is that separate mean scores are estimated for both members of an antonym pair.

*almost* and *fully* is to be expected. Third and most importantly, whether survey adjectives can be classified into closed scale/absolute and open scale/relative types independent of context will be assessed. If so, it is to be expected that a combination of a specific adjective and a specific degree modifier shows a stable rating across the 10 contexts in which that combination occurs. This means the contextual variance is expected to be small. In addition, it is to be expected that there are words that, independent of context, clearly do or do not allow modification by maximizers and approximators. Hence we expect the word variance to be large.

### 3. Results

#### 3.1. Prerequisite: the reliability of acceptability ratings of experts and non-experts

Agreement between raters is a prerequisite for further research on the linguistic behavior of gradable adjectives and degree modifiers on the basis of acceptability ratings. Recall that two groups of raters were created, based on language expertise. Hence, for both experts and non-experts, the reliability of the ratings is expected to be high.

Results show that both in the group of experts and in the group of non-experts, the reliability is .72. This means that for both groups of raters there is agreement on which combinations of adjectives and degree modifiers are acceptable and which are unacceptable. The identical reliabilities for experts and non-experts might suggest that expert and non-expert ratings are always equally reliable. However, as the reliability is a function of the number of raters (cf. Lord and Novick, 1968), the reliability is expected to decrease when fewer raters are used. Recall that 60 experts and 113 non-experts participated in this study. This means that if the juries of experts and non-experts were equally large, experts would provide more reliable ratings.

Considering the differences in the reliability of expert and non-expert ratings, it is interesting to further examine parallels and discrepancies between these two groups. The estimated eight mean scores and their variance components are shown in Table 1. The mean answers for experts and non-experts can be compared in a contrast test, which yields a  $\chi^2$ -distributed test statistic (Goldstein, 2003; Snijders and Bosker, 1999).

**Table 1**

Parameter estimates for experts and non-experts.

Item judged	$\bar{X}$	$S^2$ words <sup>a</sup>	$S^2$ persons	$S^2$ contexts	$S^2$ error
	<i>Experts (n = 60)</i>				
<i>Almost</i> + positive pole	2.80	0.11	0.41	1.00	0.39
<i>Almost</i> + negative pole	2.64		0.54	1.04	0.41
<i>Fully</i> + positive pole	3.56		0.26	1.12	0.44
<i>Fully</i> + negative pole	3.32		0.39	1.09	0.42
	<i>Non-experts (n = 113)</i>				
<i>Almost</i> + positive pole	2.67	0.11	0.45	0.86	0.33
<i>Almost</i> + negative pole	2.36		0.30	0.96	0.37
<i>Fully</i> + positive pole	3.33		0.36	0.99	0.38
<i>Fully</i> + negative pole	3.00		0.33	1.08	0.42

Note: The scale on which respondents expressed their judgment ranges from 1 (the sentence is unacceptable) to 5 (the sentence is acceptable).

<sup>a</sup> The word variance is constraint equal for all means.

Experts, on average, rate statements with a combination of an adjective and a degree modifier as more acceptable than non-experts ( $\chi^2 = 8.29$ ;  $df = 1$ ;  $p = .004$ ). This tendency was observed irrespective of whether the statements were formulated with a positive or a negative member of a word pair, and irrespective of whether the adjective in the statement is modified by *almost* or *fully*. Experts are thus more lenient than non-experts.

Besides this difference between experts and non-experts, there are also two parallels in the ratings of both types of judges. First, on average, both groups of raters rate statements with the positive member of a word pair as more acceptable than statements with the negative member of a word pair ( $\chi^2 = 12.73$ ;  $df = 1$ ;  $p < .001$ ). This effect also emerged in our pretest in which expert raters rated the same sentences as in the main experiment, but now without the degree modifiers modifying the adjectives. Therefore, this result probably indicates that, in general, sentences with a positive adjective are considered to be more acceptable than statements with a negative adjective.

Second, on average, both experts and non-experts rate sentences in which an adjective is modified by *fully*, as more acceptable than sentences in which an adjective is modified by *almost* ( $\chi^2 = 16.91$ ;  $df = 1$ ;  $p < .001$ ). This means that if antonyms are grouped into closed scale/absolute and open scale/relative types on the basis of their absolute mean rating scores in this experiment, more antonyms fall in the relative category on the basis of the empirical probe *almost*.

#### 3.2. Convergent validity: are the same adjectives acceptable with *almost* and *fully*?

If both modifiers are valid empirical probes, all sentences in which an adjective felicitously combines with *almost*, should also be considered acceptable when the adjective is combined with *fully* instead of *almost* in the same sentence. Correspondingly, all statements that are considered to be unacceptable when the adjective is modified by *almost*, should also be unacceptable when the adjective is modified by *fully*. Whether or not such convergent validity

exists, will be examined by assessing the correlation between the acceptability of *almost* and *fully*. A measure that is frequently used to classify the strength of a correlation is the proportion of shared variance. If both modifiers only combine with maximum standard adjectives, we expect the proportion of shared variance to be medium or large in size.<sup>9</sup> Our statistical model allows for testing this prediction four times: for both positive and negative adjectives rated by experts and non-experts.

Table 2 shows the correlations of the *almost* and *fully* items. For positive adjectives, there is a significant positive correlation between adjectives that combine with *almost* and *fully* ( $r = .32$ ;  $p < .05$  for experts and  $r = .65$ ;  $p < .01$  for non-experts, both are printed bold in Table 2). The proportion of overlap between *almost* and *fully* can be classified as large (42%) for the non-experts. For experts, the proportion of shared variance can be classified as medium (10%).

For negative adjectives, rated by experts, only random overlap exists between adjectives that allow modification by *almost* and *fully* ( $r = .27$ ;  $p = .44$ ). For non-experts we did observe a significant correlation ( $r = .33$ ;  $p < .05$ ). The overlap between the ratings can be classified as medium according to Cohen's criteria (11%).

**Table 2**

Correlations between the ratings for *almost* and *fully*.

Item judged	Modifier <i>Almost</i>		Modifier <i>Fully</i>	
	Positive pole	Negative pole	Positive pole	Negative pole
	<i>Non-experts</i>			
<i>Almost</i> + positive pole	–		.	
<i>Almost</i> + negative pole	<b>.85**</b>	–		
<i>Fully</i> + positive pole	<b>.65**</b>	<b>.44*</b>	–	
<i>Fully</i> + negative pole	<b>.39*</b>	<b>.33*</b>	<b>.83**</b>	–
	<i>Experts</i>			
<i>Almost</i> + positive pole	–			
<i>Almost</i> + negative pole	<b>.93**</b>	–		
<i>Fully</i> + positive pole	<b>.32*</b>	<b>.09**</b>	–	
<i>Fully</i> + negative pole	.16	<b>.27</b>	<b>.81**</b>	–

\*  $p < .05$ .

\*\*  $p < .01$ .

All in all, results show that there is considerable convergent acceptability: correlations between *almost* and *fully* are usually significant and the proportion of shared variance is medium or large in size. However, the fact the proportion of shared variance is not always medium or large in size, also indicates that there are cases in which an adjective only combines with either of the two degree modifiers.

### 3.3. The role of context in acceptability ratings

The main goal of this research is to examine if survey adjectives can – independent of context – be classified into closed scale/absolute versus open scale/relative types. If context plays only a minor role in acceptability judgments, it should be possible to distinguish adjectives that do, and adjectives that do not allow modification by *almost* and *fully*. This means that the word variance is expected to be large. Furthermore, if context plays only a minor role, a combination of an adjective and a degree modifier is expected to show a stable rating across the 10 different contexts in which that combination occurs. Consequently, the contextual variance is expected to be small.

To classify the variances (see Table 1), three 80% confidence intervals were constructed. These confidence intervals indicate the range of scores in which 80% of all words, persons, and contexts respectively lie. The word variance and the contextual variance are compared to the personal variance to determine whether these variances are relatively small or large. Hence, the word variance is expected to be larger than the person variance, but the contextual variance is expected to be smaller than the person variance.

Please note that the structure of our statistical model allows for testing this hypothesis eight times; for each estimated mean rating score this hypothesis can be tested. As the results for all eight means converge nicely (see Table 1), only the results for expert ratings on positive adjectives modified by *almost* will be discussed extensively; results for this one mean rating score give a clear indication of the general pattern found for all mean scores. Fig. 1 shows how the mean score of all positive adjectives modified by *almost* and rated by experts, varies due to word characteristics, person characteristics, and context characteristics.

As can be seen in Fig. 1, the word variance is small (confidence interval ranging from 2.38 to 3.22) compared to the person variance (confidence interval ranging from 1.98 to 3.62). So, contrary to expectation, the individual rater is more important than the specific adjective that is rated. When the word variance is interpreted in terms of the labels given to the scale, 80% of the combinations of positive adjectives and *almost* are rated as somewhere between slightly unacceptable and neither

<sup>9</sup> The proportion of shared variance can be obtained by squaring the correlation coefficient for two variables. Subsequently, the following criteria are prescribed for classifying this measure (Cohen, 1977): a correlation can be classified as *small* when the proportion of shared variance is  $<9\%$ , a correlation is *medium* when the proportion of shared variance lies between 10% and 25%, and a correlation can be called *large* when the proportion of shared variance is  $>25\%$ .

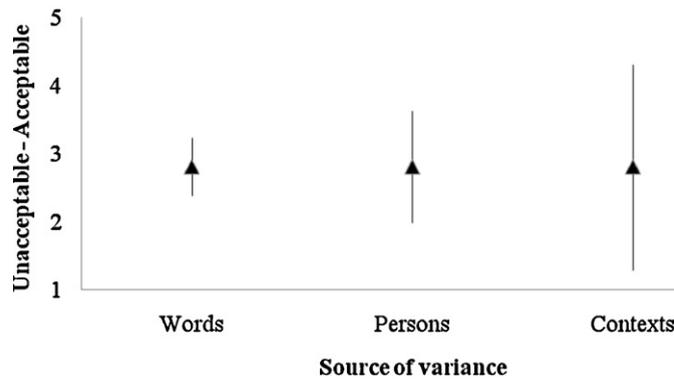


Fig. 1. Mean scores for the items judged by experts in which a positive member of a word pair has been modified by *almost*. The vertical lines indicate the 80% confidence intervals.

unacceptable nor acceptable, because only scores slightly below 2.5 and scores between 2.5 and 3.5 lie within the confidence interval. Considering these small differences between words, it is very difficult, if not impossible, to differentiate between closed scale/absolute and open scale/relative adjectives on the basis of *almost*.

Fig. 1 also shows that if a positive adjective is modified by *almost*, the ratings of a specific adjective in 10 different contexts will vary about 3 scalar points; in this case, 80% of the estimated ratings lie within the interval from 1.29 to 4.31. The contextual variance is large compared to the extent to which raters agree on these items, which varies between 1.98 and 3.62 on the scale. Hence, the specific statement in which a certain combination of an adjective and a degree modifier occurs, is more important than which rater judges the statement. When the contextual variance is interpreted in terms of the labels on the scale, results show that in one context a specific positive adjective is unacceptable when modified by *almost*, because the estimated ratings lie below 2.5. In another context that same combination is rated as not acceptable and not unacceptable, because scores between 2.5 and 3.5 lie within the confidence interval. Finally, in some contexts, the same combination of a specific positive adjective and *almost* is rated as acceptable, because scores above 3.5 also lie within the confidence interval. To sum up, the same combination of an adjective and a degree modifier is rated differently in different contexts.

To make these results more tangible, we will discuss an example of contextual variation for a specific word pair. Fig. 2 illustrates the contextual variance for the combination *almost easy* and *almost difficult* across 10 contexts. Note that Fig. 2 illustrates pure contextual variance; differences between raters are filtered out. For the exact contexts in which the combinations *almost easy* and *almost difficult* occurred, we refer to Appendix B.

Fig. 2 indicates that the acceptability scores for the combinations *almost easy* and *almost difficult* are considered to be unacceptable in context 1 and in context 2 (respectively: 'In general, law texts are easy/difficult to understand' and 'It is easy/difficult to broaden my knowledge about the government's smoking policy'). The same holds for contexts 3, 6, 8, 9, and 10. Hence, even though the extent to which these individual sentences are unacceptable varies slightly, the overall picture is clear: *easy/difficult* is associated with an open scale structure. Despite this general pattern, huge exceptions occur. The combinations *almost easy* and *almost difficult* are rated somewhere between acceptable and neither acceptable nor unacceptable in context 4

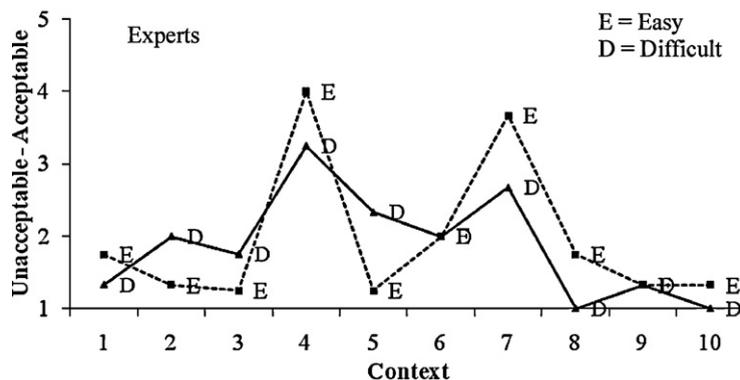


Fig. 2. Example of contextual variance: *almost easy* (E) and *almost difficult* (D) in 10 different contexts according to experts.

and in context 7 (respectively 'In the Netherlands, it is easy/difficult to buy cigarettes when you are under sixteen' and 'In the Netherlands, it is easy/difficult to buy marihuana when you are under sixteen'). Furthermore, the combination *almost difficult* is also considered to be relatively acceptable in context 5 ('I think that the number of people dying of lung cancer each year can be reduced easily/with difficulty').

Of course the results for *easy/difficult* provide just one example of contextual variability observed for two specific adjectives modified by *almost*. However, comparable results are obtained for all 19 other word pairs, that is, all 38 other adjectives examined in this study. In addition, the contextual variability is not only large for adjectives modified by *almost*; comparable results were obtained for *fully*. All in all, this means that contextual variability is ubiquitous for survey adjectives.

#### 4. Conclusion and discussion

In surveys, respondents are more likely to disagree with negative questions or statements (*This text is bad. Yes/No*) than to agree with positive ones (*This text is good. Yes/No*). Question wording, however, has shown to be much more important for one word pair than for the other (e.g. Kamoen et al., 2007). In the current study, we hypothesized that these differences may be related to a natural classification of adjectives based on the scale structure (open/closed) and the type of comparison standard (relative/absolute). A prerequisite for testing this hypothesis is that survey adjectives can be classified unambiguously into open scale/relative and closed scale/absolute types. The current research was set up to examine whether such a classification can be made.

In a judgment experiment, two groups of language users (experts and non-experts) rated the acceptability of sentences in which a survey adjective was modified by the maximizer *fully* or the approximator *almost*. Both types of degree modifiers can be used as an empirical probe to determine the scale structure and the type of comparison standard: closed scale/absolute adjectives should allow modification by these modifiers, whereas the use of open scale/relative adjectives should be unacceptable when combined with these modifiers (e.g. Kennedy and McNally, 2005; Kennedy, 2007; Rotstein and Winter, 2004). The same adjective and degree modifier combination was examined in a range of 10 sentences to determine if the adjective could be classified relatively unambiguously, that is, relatively independent of context.

Results show that the two groups of judges agree on which combinations of adjectives and degree modifiers are acceptable and which combinations are unacceptable. Also in line with our expectations, the two empirical probes show considerable convergent validity: in one and the same context, an adjective generally combines both with *fully* and with *almost*, or fails to combine with *fully* and with *almost*. Nevertheless, results show that a classification of adjectives cannot be made unambiguously. The first reason is that the same combination of a survey adjective and a degree modifier is rated differently across the 10 contexts in which that combination occurs. Second, independent of context, it is not possible to differentiate between survey adjectives that clearly do and survey adjectives that clearly do not allow modification by the two degree modifiers. For survey practice, this means that it is impossible to distinguish between word pair variation in response effects for contrastive questions using this typology.

In the current study we applied a semantic classification of adjectives to a set of twenty antonyms often used in surveys. Of course, semantic theory is not designed for classifying survey adjectives and distinguishing between word pair variation in survey framing effects. Therefore, in some sense, there is a gap between the theoretical study of semantics and the application of a semantic classification to the very specific language use context of attitude surveys. Yet, it is worthwhile that survey researchers try to apply semantics to their field of study, especially because the field of semantics may also benefit from such attempts. In our view, the current study contributes to semantics in at least two ways.

First, this study contributes to semantics in a methodological respect: it shows how informant ratings can be applied to examine linguistic behavior. Compared to corpus research, which is a more frequently used method for examining linguistic behavior, this research method has an important advantage. Corpus analyses provide valuable information about the linguistic behavior of adjectives and degree modifiers, but a drawback of this method is that it only provides positive instances of an adjective's use. Therefore, corpus data cannot show which linguistic combinations of adjectives and degree modifiers are unacceptable. Language users' judgments do provide information about the unacceptability of adjective and degree modifier combinations. Furthermore, an advantage of applying informant ratings on a large scale – instead of using only one informant or a small group of informants – is that the reliability of the ratings can be examined. Reliability is a reflection of agreement, and if there is no agreement on which combinations of adjectives and degree modifiers are acceptable, there is in fact no linguistic behavior to describe.

Second, the current study is relevant for semantics because it raises the question as to why, counter to expectation, so much contextual variation was observed in the linguistic behavior of adjective and degree modifier combinations for the stimuli used here. We can only speculate about the possible causes for this observation.

One possibility is that there is contextual variation because there are contexts in which an adjective and degree modifier combination is judged to be acceptable, even though no real maximum degree reading of the adjective is involved. This "false alarm" explanation may be further specified in at least two ways. On the one hand, the modifiers may have other preferences than maximum standard adjectives. On the other hand, there may be situations in which the modifier, although syntactically in construction with the adjective, takes scope over a different part of the sentence

and there finds a closed scale to work over. An example of this latter explanation has been given in Kennedy and McNally (2005:366). In their example 53.b, they show that the sentence *The baby's face is completely hot* is acceptable because *completely* takes scope over the subject of the sentence (*the baby's face*) and therein finds a closed scale to work over, resulting in something close in meaning to *all of the baby's face is hot*. Unfortunately, it is impossible to further examine these “false alarm explanations” based on the data obtained in the current study. As the interpretation of test sentences has not been measured, it is impossible to further examine the scope of the modifiers. Post hoc specifications of potential alternative combinatorial preferences of the two degree modifiers are difficult, because each adjective and degree modifier combination occurred in a set of random contexts and therefore sentence characteristics may be confounded.

Although it is important to investigate potential “false alarm” situations in a future study, it should also be stressed that such explanations are unlikely to account for all the contextual variance observed in the current study. Our results show that there is large convergence in adjectives that combine with *almost* and *fully* in one and the same sentence. This suggests that there are cases in which for both *almost* and *fully* deviant behavior is exposed in the same context. Hence, only if the two degree modifiers largely reflect the same “alternative combinatorial preferences” or the same “alternative scopes” all contextual variation can be accounted for.

Other than the false alarm explanation, the fact that the same adjective and degree modifier combination is acceptable in one context and unacceptable in another context may be caused by actual contextual variation in the scale structure and the type of comparison standard of adjectives. This would suggest that one and the same adjective is associated with a closed scale and interpreted against an absolute comparison standard in one context, but associated with an open scale and interpreted against a relative comparison standard in another context. In the semantic literature, non-default interpretations of adjectives are mentioned (e.g. Bolinger, 1972; Kennedy and McNally, 2005; Kennedy, 2007). However, how often such non-default behavior is exposed, has not yet been examined. Results of the current study indicate that at least for the survey adjectives studied here, exceptions occur quite often and the default scale structure/type of comparison standard associated with each adjective is quite weak. In addition, the *easy/difficult* example illustrates that not only presumably closed scale/absolute word pairs, but also open scale/relative pairs show non-default behavior. This is contrary to expectation, as it has generally been assumed that non-default behavior of adjectives and degree modifiers arises mainly for closed scale adjectives (e.g. Kennedy and McNally, 2005).

All in all, our results support the idea that the examined semantic classifications of adjectives are not so much by default attached to an adjective (as seems to be pronounced in e.g. Kennedy and McNally, 2005; Kennedy, 2007; Rotstein and Winter, 2004), but merely construed in a specific context (compare Paradis, e.g. 1997, 2001, 2008, 2009). Future research is challenged to further model the extent to which the scale structure and the type of comparison standard are lexical characteristics of adjectives, and to what extent these are characteristics that must always be uniquely determined in a certain context. The construction of such a model will benefit greatly from research in which the linguistic behavior and the interpretation of adjectives are examined in a variety of contexts.

## Acknowledgements

The authors thank Elena Tribushinina and Louise McNally for comments on earlier versions of this article. In addition, the authors thank the two anonymous reviewers of Journal of Pragmatics for their detailed feedback.

## Appendix A. Word pairs used

Dutch word pair	English translation
Wenselijk/onwenselijk	Desirable/undesirable
Vernieuwend/ouderwets	Modern/old-fashioned
Verstandig/onverstandig	Wise/unwise
Redelijk/onredelijk	Reasonable/unreasonable
Slim/dom	Smart/stupid
Makkelijk/moelijk	Easy/difficult
Goed/slecht	Good/bad
Progressief/conservatief	Progressive/conservative
Boeiend/saai	Fascinating/boring
Interessant/oninteressant	Interesting/uninteresting
Aanvaardbaar/onaanvaardbaar	Allowable/unacceptable
Rechtvaardig/onrechtvaardig	Just/unjust
Begrijpelijk/onbegrijpelijk	Understandable/'ununderstandable'
Adequaat/inadequaat	Adequate/inadequate
Acceptabel/onacceptabel	Acceptable/unacceptable
Logisch/onlogisch	Logical/illogical
Positief/negatief	Positive/negative
Consequent/inconsequent	Consequent/inconsequent
Duidelijk/onduidelijk	Clear/unclear

## Appendix B

Example of 10 contexts for word pair *makkelijk/moeilijk* (*easy/difficult*) For each of the 10 contexts, we first provide the Dutch original sentence, then the word-by-word English translation, and lastly, a free English translation, which indicates the original communicative meaning.

1. Wetteksten zijn over het algemeen makkelijk/moeilijk begrijpbaar.  
Law.texts are over the general easy/difficult understandable.  
'In general, law texts are easy/difficult to understand.'
2. Het lijkt me makkelijk/moeilijk om me te verdiepen in het rookbeleid van de overheid.  
It seems me easy/difficult to me to lose into the smoking.policy of the government.  
'It is easy/difficult to broaden my knowledge about the government's smoking policy.'
3. Wetteksten zijn over het algemeen makkelijk/moeilijk leesbaar.  
Law.texts are over the general easy/difficult readable.  
'In general, law texts are easy/difficult to read.'
4. Het is makkelijk/moeilijk om in Nederland als zestienjarige sigaretten te kopen.  
It is easy/difficult to in Netherlands like sixteen.year.old cigarettes to buy.  
'In the Netherlands, it is easy/difficult to buy cigarettes when you are under sixteen.'
5. Ik denk dat het aantal mensen dat jaarlijks aan longkanker sterft makkelijk/moeilijk kan worden teruggedrongen.  
I think that the number people that yearly to lung.cancer dies easy/difficult can become reduced.  
'I think that the number of people dying of lung cancer each year can be reduced easily/with difficulty.'
6. De nieuwe regels ten aanzien van roken op de werkplek zijn makkelijk/moeilijk te begrijpen.  
The new rules with respect to smoking on the working place are easy/difficult to understand  
'The new rules with respect to smoking in workplaces are easy/difficult to understand.'
7. Het is makkelijk/moeilijk om in Nederland als zestienjarige wiet te kopen.  
It is easy/difficult to in Netherlands like sixteen.year.old pot to buy.  
'In the Netherlands, it is easy/difficult to buy marihuana when you are under sixteen.'
8. Ik denk dat het aantal rokers in Nederland makkelijk/moeilijk kan worden teruggedrongen.  
I think that the number smokers in Netherlands easy/difficult can become reduced.  
'I think that in the Netherlands the number people who smoke can be reduced easily/with difficulty.'
9. Het lijkt me makkelijk/moeilijk om me verder te verdiepen in politieke kwesties van Nederland.  
It seems me easy/difficult to myself further to lose to political issues of Netherlands.  
'I think it is easy/difficult to broaden my knowledge about political issues in the Netherlands.'
10. Het aantal gevallen van longkanker in Nederland kan makkelijk/moeilijk worden teruggedrongen.  
The number cases of lung.cancer in Netherlands can easy/difficult become reduced.  
'In the Netherlands, the number of lung cancer cases can be reduced easily/with difficulty.'

## Appendix C. More information about the multi-level model

In this sample, each acceptability rating is nested both within raters, and within items. Common statistical techniques, such as regression analysis, are not equipped to handle these two random factors at the same time. When using such techniques for analyzing complex data, the variance will be underestimated, and, as a consequence, the standard error will be underestimated as well. Therefore, the chance of making a type 1 error increases. Multi-level analysis can handle multiple random factors at the same time. Therefore, for complex data, this statistical technique is to be preferred over traditional analyses of variance.

In the multi-level model used for this experiment, eight mean acceptability scores are estimated, one for each combination of the degree modifier (*almost* or *fully*), the polarity of the adjective (positive or negative), and the type of rater (expert or non-expert). The three sources of variance (word variance, person variance, and contextual variance) are estimated for each mean score.

As an example, let us consider the mean acceptability score for all positive adjectives modified by *almost* and rated by experts. The actual mean rating score for this set of adjectives is of minor importance. More importantly, this model estimates how much, on average, the mean score of one specific positive adjective modified by *almost* and rated by experts,

let's say the mean score of all instances of *almost clear*, deviates from this overall mean score of all positive adjectives modified by *almost* and rated by experts. This source of variance is what we call the word variance. If there are combinations of positive adjectives modified by *almost* and rated by experts that are clearly acceptable, whereas other combinations are clearly not, the word variance will be large.

Second, the person variance indicates how, on average, the mean rating of all positive adjectives modified by *almost* and rated by one single expert rater, deviates from the overall mean rating score for all expert raters for all positive adjectives modified by *almost*. This source of variance tells us whether the overall strictness of the raters varies. Hence, by estimating the variation between different raters, we account for the fact that raters may use the response scale differently.

Third, we estimate how similar, on average, one and the same rater rates the same combination of an adjective and a degree modifier in different contexts. For example, how does rater *X* rate the combination *almost clear* in context three, as compared to context seven? This source of variance tells us to what extent the rating of the same adjective/degree modifier combination varies across contexts. Please do note that this source of variance also (necessarily) includes random error variance because it is modeled at the lowest level of the sampling hierarchy.

Now let us formalize the model. If  $Y_{i(jk)}$  is the  $i$ th ( $i = 1, 2, \dots, I_{(jk)}$ ) rating of individual  $j$  ( $j = 1, 2, \dots, J$ ) of word  $k$  ( $k = 1, 2, \dots, K$ ), and there are eight dummies ( $D$ ) one for each combination of the method (indicated by an  $A$  for *almost* or an  $F$  for *fully*), the polarity of the adjective (indicated by a  $P$  for positive and an  $N$  for negative) and the type of language user that rated the statement (indicated by an  $E$  for experts or an  $N$  for non-expert), the model can be written as in Eq. (C.1).<sup>10</sup>

Eq. (C.1): multi-level model used to predict the ratings and model the variances

$$Y_{i(jk)} = D_{(APN)}_{i(jk)}(\beta_1 + e_{1i(jk)} + u_{10(jk)}) + D_{(ANN)}_{i(jk)}(\beta_2 + e_{2i(jk)} + u_{20(jk)}) + D_{(APE)}_{ijk}(\beta_3 + e_{3i(jk)} + u_{30(jk)}) \\ + D_{(ANE)}_{i(jk)}(\beta_4 + e_{4i(jk)}) + D_{(FPN)}_{i(jk)}(\beta_5 + e_{5i(jk)} + u_{50(jk)}) + D_{(FNN)}_{i(jk)}(\beta_6 + e_{6i(jk)} + u_{60(jk)}) \\ + D_{(FPE)}_{i(jk)}(\beta_7 + e_{7i(jk)} + u_{70(jk)}) + D_{(FNE)}_{i(jk)}(\beta_8 + e_{8i(jk)} + u_{80(jk)} + v_{000k}) \quad (C.1)$$

All residuals are normally distributed with an expected value of zero, and a variance of respectively  $S^2_{ei(1, \dots, 8)}, \dots, S^2_{ui(1, \dots, 8)}, \dots, S^2_v$ . Please note that in this model the word variance ( $S^2_v$ ) is estimated only once for all eight mean scores together. This is a constraint of the model. Furthermore, the contextual variance  $S^2_{ei(1, \dots, 8)}$  consists of actual contextual variance, and of (random) error variance.

## References

- Amaral, Patricia, 2006. On the semantics of *almost*. Paper presented at the annual conference of the Linguistic society of America, Albuquerque, 5–8 January.
- Bierwisch, Manfred, 1989. The semantics of gradation. In: Bierwisch, M., Lang, E. (Eds.), *Dimensional Adjectives*. Springer-Verlag, Berlin, pp. 71–261.
- Bishop, George F., Hippler, Hans-Juergen, Schwarz, Norbert, Strack, Fritz, 1988. A comparison of response effects in self-administered and telephone surveys. In: Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls II, W.L., Waksberg, J. (Eds.), *Telephone Survey Methodology*. Wiley, New York, pp. 273–282.
- Bolinger, Dwight, 1972. *Degree Words*. The Hague, Mouton.
- Cohen, Jacob, 1977. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, New York.
- Croft, William, Cruse, Alan D., 2004. *Cognitive Linguistics*. Cambridge University Press, Cambridge.
- Cruse, Alan D., 1986. *Lexical Semantics*. Cambridge University Press, Cambridge.
- Goldstein, Harvey, 2003. *Multilevel statistical models*. Edward Arnold, London.
- Hamilton, Helen W., Deese, James, 1971. Does linguistic marking have a psychological correlate? *Journal of Verbal Learning and Verbal Behavior* 10 (6), 707–714.
- Hayes, Samuel P., 1939. The inter-relations of political attitudes: II. Consistency in voters' attitudes. *Journal of Social Psychology* 10 (3), 359–378.
- Hippler, Hans-Juergen, Schwarz, Norbert, 1986. Not forbidding isn't allowing: the cognitive basis of the forbid/allow asymmetry. *Public Opinion Quarterly* 50 (1), 87–96.
- Hoeksema, Jack, 2008. Discourse scalarity: the case of Dutch *helemaal*. Paper presented at la scalarité dans tout ses aspects. Gent, 15–26 December.
- Holleman, Bregje C., 1999. Wording effects in survey research. Using meta-analysis to explain the forbid/allow asymmetry. *Journal of Quantitative Linguistics* 6 (1), 29–40.
- Holleman, Bregje C., 2000. The forbid/allow asymmetry. On the cognitive mechanisms underlying wording effects in surveys. Doctoral Dissertation, Rodopi, Amsterdam – Atlanta.
- Kamoen, Naomi, Holleman, Bregje C., Van den Bergh, Huub, 2007. Hoe makkelijk is een niet moeilijke tekst? Een meta-analyse naar het effect van vraagformulering in tekstevaluatieonderzoek. (How easy is a text that is not difficult? A meta-analysis on wording effects in text evaluation research). *Tijdschrift voor taalbeheersing* 29 (4), 314–330.
- Kamoen, Naomi, Holleman, Bregje C., Van den Bergh, Huub. How easy is a text that is not difficult? Comparing answers to positive, negative, and bipolar questions, submitted for publication.
- Kennedy, Christopher, 1999. Projecting the adjective: the syntax and semantics of gradability and comparison. Doctoral Dissertation, Garland, New York.
- Kennedy, Christopher, 2001. Polar opposition and the ontology of 'degrees'. *Linguistics and Philosophy* 24 (1), 33–70.
- Kennedy, Christopher, 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30 (1), 1–45.
- Kennedy, Christopher, McNally, Louise, 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81 (2), 345–381.
- Krosnick, Jon A., Schuman, Howard, 1988. Attitude, intensity, importance, and centrality and susceptibility to response effects. *Journal of Personality and Social Psychology* 54 (6), 940–952.
- Levin, Irwin P., Gaeth, Gary J., 1988. Framing of attribute information before and after consuming the product. *Journal of Consumer Research* 15 (3), 374–378.
- Levin, Irwin P., Schneider, Sandra L., Gaeth, Gary J., 1998. All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes* 76 (2), 149–188.
- Lord, Frederic M., Novick, Melvin R., 1968. *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.
- Meyerowitz, Beth E., Chaiken, Shelly, 1987. The effect of message framing on breast self-examination attitudes, intentions and behavior. *Journal of Personality and Social Psychology* 52 (3), 500–510.

<sup>10</sup> Please note that in fact a cross-classified model is in operation, because the ratings are nested within both individuals and words (cf. Quené and Van den Bergh, 2008).

- Molenaar, Nico J., 1982. Response-effects of “formal” characteristics of questions. In: Dijkstra, W., Van der Zouwen, J. (Eds.), *Response Behaviour in the Survey-Interview*. Academic Press, London, pp. 49–89.
- Paradis, Carita, 1997. Degree modifiers of adjectives in spoken British English. Doctoral Dissertation, Lund University Press, Lund.
- Paradis, Carita, 2000. It's well weird: degree modifiers of adjectives revisited: the nineties. In: Kirk, J. (Ed.), *Corpora Galore: Analyses and Techniques in Describing English: Papers from the Nineteenth International Conference on English*. Rodopi, Amsterdam, pp. 147–160.
- Paradis, Carita, 2001. Adjectives and boundedness. *Cognitive Linguistics* 12 (1), 47–65.
- Paradis, Carita, 2008. Configurations, construals and change: expressions of DEGREE. *English Language and Linguistics* 12 (2), 317–343.
- Paradis, Carita, Willners, Caroline, 2006. Antonymy and negation – the boundedness hypothesis. *Journal of Pragmatics* 38 (7), 1051–1080.
- Quené, Hugo, Van den Bergh, Huub, 2004. On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication* 43 (1–2), 103–121.
- Quené, Hugo, Van den Bergh, Huub, 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59 (4), 413–442.
- Rotstein, Carmen, Winter, Yoad, 2004. Total adjectives vs. partial adjectives: scale structure and higher-order modifiers. *Natural Language Semantics* 12 (3), 259–288.
- Rugg, Donald, 1941. Experiments in wording questions II. *Public Opinion Quarterly* 5 (1), 91–92.
- Schuman, Howard, Presser, Stanley, 1981. Questions and Answers in Attitude Surveys. Experiments on Question form Wording and Context. Academic Press, London.
- Snijders, Tom A.B., Bosker, Roel J., 1999. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage Publishers, London.
- Syrett, Kristen L., 2007. Learning about the structure of scales: adverbial modification and the acquisition of the semantics of gradable adjectives. Doctoral Dissertation, Northwestern University, Chicago.
- Tribushinina, Elena, 2008. Cognitive reference points: semantics beyond the prototypes in adjectives of space and colour. Doctoral Dissertation, LOT, Utrecht.
- Tribushinina, Elena, 2009. The linguistics of zero: a cognitive reference point or a phantom? *Folia Linguistica* 43 (2), 417–466.
- Vanden Wyngaerd, Guido, 2001. Measuring events. *Language* 77 (1), 61–90.
- Waterplas, Lina, Billiet, Jaak, Loosveldt, Geert, 1988. De verbieden versus niet toelaten asymmetrie. Een stabiel formuleringseffect in survey-onderzoek? (The forbid/allow asymmetry. A stable response effect in survey research?). *Mens en Maatschappij* 63, 399–415.

**Naomi Kamoen** is a PhD-student at the Utrecht Institute of Linguistics OTS at Utrecht University. Her research focuses on language use and cognitive processes of text interpretation, specifically targeted at processing contrast relations in a survey situation and the semantics of contrast relations.

**Bregje Holleman** is an assistant professor at the Department of Dutch and the Utrecht Institute of Linguistics OTS at Utrecht University. The focus of her research and teaching is on language use and cognitive processes of text interpretation. Her main interests are framing effects of positive versus negative wordings as they occur in survey questions and persuasive texts.

**Rick Nouwen** is a senior research fellow at the Utrecht Institute for Linguistics OTS. His research focuses on the semantics-pragmatics interface. In the past, he worked on quantification and scalarity. Currently, he is leading the research project *Degrees under discussion*, funded by the Dutch organization for scientific research (NWO).

**Ted Sanders** is a professor of Discourse Studies and Dutch language use at Utrecht University. His research focuses on discourse structure and cognitive processes of text interpretation. Currently, he heads the research project *Causality and Subjectivity in Discourse and Cognition*, funded by the Dutch organization for scientific research (NWO).

**Huub van den Bergh** is a professor of teaching and testing of language proficiency at Utrecht University. His research is on cognitive mechanisms of text interpretation and text production. The focus of his research is on writing research and methodological features of measuring processes of text interpretation and production.