Numerical Discrepancies Between 'Some' and 'A Few' A Basis for Dutch Scalar Implicature Research

Rien Debrouwer (Rien.Debrouwer@gmail.com)
Department of Psychology, University of Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Walter Schaeken (Walter.Schaeken@ppw.kuleuven.be)
Department of Psychology, University of Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Abstract

Horn scales are a popular vehicle in the investigation of implicatures. Yet even this most user-friendly of implicature research categories is plagued by methodological and extrapolating difficulties. One of these difficulties is the possible existence of pungent semantic discrepancies that get lost in translation. To form a basis for past and future Dutch scalar implicature research, we investigated the popular quantifier 'some'. In an experiment we registered different elements that make up its numerical description (i.e. minimal, most likely and maximal value) and compared them to those of other quantifiers from its Horn scale. The experiment showed that the parameter values for some are overall higher than those for a few. A scaling effect on some, however, appears to blur some's discrepancies with a few for lower population sizes.

Keywords: some; a few; scalar implicatures; numerical description; population size.

Introduction

Communication consists of a process where: a sender encodes his or her message into language, transmits the message through a certain medium or channel (e.g., speech), and a receiver decodes and interprets the meaning of that message (Levelt, 1989). A message does not merely consist of the semantic value of a series of lexemes, ordered according to grammatical convention (i.e. the *linguistic meaning* of the message). Most often, the receiver has to attempt interpreting the *pragmatic meaning* of the message as well, i.e. the so-called implied subtext with all its relevant connotations. The message in (1), for example, is likely not actually an inquiry on someone's aptitude at causing fissures in windows. It refers to any nearby window, the context suggests a physical window (i.e. not a metaphorical one), the used syntax (i.e. "Could you …") is common practice for conveying a request for active behavior, in the current context the word "crack" is likely to be interpreted as the American English slang word for "open", and the messenger's posture and gestures might suggest that the goal for this request is to lower the indoor temperature. In (2), the linguistic message may be at odds with the pragmatic one as well. The receiver has to inspect the messenger's voice intonation, gestures, the general context and former experience in order to discern whether this message was meant: ironically, sarcastically, in gest, social protocol, heartfelt, … .

(1) Could you crack a window?
(2) It is nice to meet you.

Such information is not explicitly mentioned, though vital for proper communication. It is part of the conventions that make language more compact and manageable. It is not practical and opportune to repeat all this information with every conveyed message: due to its sheer magnitude, and because human beings tend to think faster than they can articulate (Levinson, 2000). In attempting to resolve this laryngeal bottleneck, human language has developed certain pragmatic, culturally defined conventions or rules that provide linguistic messages with implied information for the receiver to infer. Grice (1989) introduced the name *implicatures* for these pragmatic, implied meanings of messages. He discerned between Conventional Implicatures and Conversational Implicatures. Example (3) illustrates the former. The linguistic meaning of (3) is that someone named Paul had feelings of tiredness and satisfaction. An implied message in (3) states that, out of its several lexical meanings (e.g., merely, yet, in spite of), the word 'but' is to be interpreted as 'in spite of'. This pragmatic meaning of the message stems from general linguistic convention on how 'but' is to be interpreted given the grammatical build of the sentence.

(3) Paul felt tired, but satisfied.

Conversational Implicatures, on the other hand, cannot be derived from such a secluded inspection of a sentence. Looking back at (1), the pragmatic meaning of "Could you crack a window" is part of linguistic convention, i.e. a Conventional Implicature. The reason for this request could only be detected from conversational cues (e.g., body language indicating feeling cold), i.e. a Conversational Implicature. While Conventional Implicatures rely on convention, Conversational Implicatures depend on certain rules of conversation known as Grice's four maxims (Grice, 1989). The messenger has to abide by these maxims for the pragmatic meaning of a message to get across properly, otherwise Conversational Inferences cannot be correctly made. Hence, conversational messages have to be:

- Informative (Maxim of Quantity),
- Truthful (Maxim of Quality),
- Relevant to the conversation subject (Maxim of Relation),
- And appropriately delivered (Maxim of Manner) to avoid for example ambiguity.

Example (2) is informative if the correct intensity is phrased (i.e., nice, not great or OK), truthful if the speaker means it, relevant if said at an introduction, and appropriately delivered if spoken with a sincere demeanor. These four conditions being met, the listener may make the presumed correct pragmatic inference that the speaker indeed finds it nice to meet them (in this case being the same as the linguistic message). If, for example, the delivery was in a sarcastic tone, a different pragmatic meaning might be assigned to the message.

Conversational Implicatures can be subdivided into Generalized Conversational Implicatures and Particularized Conversational Implicatures. The difference between the two boils down to the level in which they depend on contextual factors. Particularized Conversational Implicatures are particular to a specific conversation. In (4), Tom's utterance may hold an implied message, a pragmatic meaning that Amy mistook Paul's tiredness for looking unsatisfied. This implicature cannot be drawn from Tom's utterance itself, only from the broader conversational context. If Tom had said: "He was tired." in response to a different question, for example: "Why did the hare take a nap midrace?", the aforementioned implicature would not have been part of the pragmatic meaning of his message.

(4) Amy: "Did Paul seem unsatisfied to you?"
Tom: "He was tired."

Generalized Conversational Implicatures, such as in (2), can be derived from the message itself (including delivery and body language). A specific type of Generalized Conversational Implicatures is Scalar Implicatures. Implicatures of the scalar kind are relatively clear-cut and lenient to manipulation, with a relatively low chance at confounding variables. Conventional Implicatures are conceptually more difficult to discern from the linguistic meaning of messages than Conversational Implicatures are, and Particular Conversational Implicatures' higher dependence on context factors makes them a lot harder to control compared to Generalized Conversational Implicatures. Of this latter category, Scalar Implicatures are the best known and most explored. Therefore this scalar type of implicatures is a welcome and often preferred subject of research. Previous research shows for example that making a pragmatic inference is not the default behavior, even though certain scalar inferences are found to be made in high percentages of cases (e.g., cf. Table 1). Children tend to interpret messages more as their linguistic meaning than as their pragmatic meaning (Noveck, 2001). They respond more pragmatically as this behavior is more saliently indicated to be the goal of the task (Guasti et al., 2005), yet even then not as often as adults do (Guasti et al., 2005; Papafragou & Musolino, 2003). Children may also generate significantly more scalar inferences for one quantifier (e.g., a few) compared to another (e.g., some) where this difference disappears towards adulthood (Pouscoulous et al., 2007). One plausible explanation for this difference between children and adults in interpreting implicative messages is that making scalar inferences requires mental processing. Children don't have as much of these mental resources, resulting in fewer implicatures being produced, and even fewer for more complex quantifiers (Pouscoulous et al., 2007). Yet, adults also require additional time (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Noveck & Posada, 2003), working memory (De Neys & Schaeken, 2007; Dieussaert et al., 2011) and other cognitive resources (Dieussaert et al., 2011) to process the pragmatic meaning of messages. If more of these mental resources are otherwise engaged, fewer inferences will be made.

These findings are not only valuable in exploring the inner workings of implicature processing, they also have repercussions for the paradigms used in research on scalar inferences. Several methodological features influence the frequency of implicature generation. Next to aforementioned effect of task structure (e.g., dual task paradigm with adults), and salience of the goal of the task (seven-year-olds), in younger children (five- but not seven-year-olds) the type of task is paramount. Action-Based Tasks, for example, stimulate far more production of scalar inferences in five-year-olds than Truth-Value Judgement Tasks do (Janssens & Schaeken, 2012). The content of the message that is to be interpreted, is vital as well. More semantically complex quantifiers (cf. supra) or more abstract statements (Janssens & Schaeken, 2012) result in fewer pragmatic interpretations in respectively nine- and seven-year-olds, and the specific syntax of the statement influences implicature production in adults (Breheny, et al., 2006). Even though scales of quantifiers are a very popular representation of implicatures, their interpretation appears to be prone to task- and procedure-related influences (e.g., training: Papafragou & Musolino, 2003). Therefore one has to take great care in considering such paradigm discrepancies when comparing experiments and generalizing results. Moreover, these extrapolation issues emphasize the importance of a strong basis, a solid central concept for paradigms in implicature research. Yet, as we will explore next, even a vehicle as straight-forward as scales of quantifiers could do with a more uniform understanding.

The scales used in scalar implicatures are called Horn scales, named after Laurence R. Horn who first introduced them (Horn, 1972). Geurts defines Horn scales as: ”(…) simply a sequence of increasingly informative expressions.” (Geurts, 2010). These expressions are all part of the same variable or dimension. Take, for example, a grouping of quantifiers that all indicate a certain degree of temperature: cold/cool/warm/hot. Such quantifiers do not represent fixed measures of temperature, yet they can easily be ranked on a temperature-related dimension line (i.e. hotness, or coldness). In research on scalar implicatures, a weaker, logically less informative term is compared to a stronger, logically more informative term (e.g., warm and hot). In their logical semantic meaning, the stronger term includes the weaker term. This less informative, weaker term refers to a section of the measurement the stronger term represents: if someone has five apples they also have three apples, if it is hot outside it is also warm outside. Yet pragmatically, these might seem like incorrect claims, due to Grice’s Maxim of Quantity. The interpretation that, given a stronger term, the weaker term is incorrect, is called a scalar implicature or scalar inference.

It is even ill-advised to casually compare results of different studies if they did not implement the same Horn scale. Let us consider for example: Noveck (2001, Exp. 1) who registered implicatures for 65% of participants on the might/must scale, Pijnacker et al. (2009) with 54% for the scale or/and, Papafragou and Musolino (2003, Exp. 1) who found 93% for start/finish and 100% for the numeral scale two/three (where three counted all members of the group). Studies can have very different results in using the same Horn scale, due to intended manipulations or methodological influences. For some/all, the most popular scale, Papafragou and Musolino (2003, Exp. 1) and Zevakhina (2012) found 93% of implicature generation. Yet other results were found for this scale reading for example 59% (Bott & Noveck, 2004, Exp. 3; Noveck, 2001, Exp. 3), or even down to 34% (Geurts & Pouscoulous, 2009, Exp. 2).

The or/and scale, with aforementioned result of 54% (Pijnacker et al., 2009), brought results of 25% in a different study (Chevallier et al., 2008, Exp.1). For other scales, similar fluctuations in results can be presented. Undoubtedly these differences in results are mostly due to the experimental manipulations of the specific studies. But it does pose questions on how to validly extrapolate from individual studies and formulate funded, meaningful statements regarding the workings of scalar implicatures in general, i.e. regardless of which Horn scales were used. Most research implements the some/all scale to test a claim regarding scalar implicatures without taking into account the existence of many other Horn scales and other quantifiers within a Horn scale that could produce significantly different results. Marty, Chemla and Spector (2013) illustrated this concern by investigated aforementioned influence of working memory strain, both with a some/all scale as with numerals. In their experiment using numerals, a higher workload was contradictorily accompanied with a higher preference of the pragmatic meaning.

In order to gain some uniformity between studies, despite experimental differences between their paradigms, critical quantifiers in studies should be identified on a uniform measure. For existential quantifiers, we suggest using worldly categories with fixed population counts. Some/all could for example be expressed as there being 83 cars (i.e. all = 83) at a certain location, and participants could be asked to define ‘some’ as an amount of those cars. The current study looks into such a numerical definition for the quantification pair some/all.

Pouscoulous et al. (2007) raised an important point in that research on scalar implicatures is done by different research groups in different countries, i.e. in different languages, and therefore may exhibit small lexical differences. The comparison of some versus all, for example, has been researched in a number of languages, using the translation of these quantifiers from English to the language in question. A non-exhaustive list of languages, in which some/all implicatures were investigated, could be: Dutch (Belgium e.g., De Neys & Schaeken, 2007; the Netherlands e.g., Geurts & Pouscoulous, 2009, Exp. 2), English (e.g., Katsos & Bishop, 2011), French (e.g., Bott & Noveck, 2004), German (e.g., Röhrig, 2010), Greek (e.g., Breheny, Katsos, & Williams, 2006), Italian (e.g., Guasti et al., 2005). Previous studies have shown distinct differences in semantic nuances dependent on the language(s) one is brought up with (e.g., Dutch versus French versus bilingually Dutch and French: Ameel et al., 2004). Concerning implicature research, Pouscoulous et al. (2007) specified the issue in French experiments to the quantifiers ‘quelques’ and ‘certains’ both being valid translations of ‘some’. The use of certains produced fewer scalar inferences in 9-year-old children than quelques did, plausibly due to the former being of a higher lexical complexity (by adding a partitive attribute). In Dutch, analogue to French, ‘some’ can be translated as ‘sommige’ or as ‘enkele’ (Van Dale, 2014). Therefore a similar investigation should be held on the semantic differences between sommige and enkele, to improve the interlingual extrapolation of research using these translations in its paradigm.

This study aims to be a starting point for that semantical comparison between sommige (to improve readability, from here on identified as ‘some’) and enkele (henceforth ‘a few’). We will look into their numerical description, i.e. a numerical expression of their position on the none/some/all Horn scale used in implicature studies. In a renowned Dutch dictionary (Van Dale, 2014), both some and a few are described as being a low amount. Yet, in comparison, a few is more often described as referring to one single unit. Therefore, and intuitively, we hypothesize that in general a few indicates a lower amount than some.

We will enquire about the preferred value, i.c. the most likely amount the quantifiers indicate given a certain population size. As elaboration on this numerical estimate, the minimal and maximal value the quantifier could represent are requested as well, for several population sizes and categories. As a control for our method, the quantifier 'most' will be added to the inquiry: most is likely to be considered a more informative quantifier (i.e. representing a higher amount) than a few and some. The partitive attribute of most is also clearer than that of a few or some: most is named after indicating over half of the population. This partitive feature is not the core of the current study, but given its proclaimed central role in the semantic difference between the French analogues quelques and certains (Pouscoulous et al., 2007), we included it in our investigation. This partitive quality might translate into the parameter values for some being more scaled to the population size than those for a few. Our quantifiers will be manipulated between subjects, in order to avoid that participants' responses would be influenced by the presentation of other quantifiers than the one at hand. We used a number of non-linear population sizes in order to avoid that every quantifier would be assessed proportionate-by-default to the previous population size (cf. Borges & Sawyers, 1974).

Method

Participants

216 first-year bachelor students in psychology participated in partial fulfilment of course requirements (17-28 years of age, M = 18.5; female: 168, male: 48). All participants were native Dutch speakers.

Design, Material and Procedure

Each participant was presented a *pen-and-paper questionnaire* in Dutch. It consisted of three items. They were constructed in the same general fashion.

The instructions for each item started off with: 'Imagine that at a certain location there are <amount> <category>.' The first item spoke of 1019 flowers, the second item regarded 10 chairs and the third item 83 cars. The amounts and categories were matched so that they would make sense, so that participants might be able to envision in a lifelike situation for them. In Belgium it is a tradition in certain folk festivals to display a huge flower tapestry on the floor of a big square. Such a scene, or for example a vast meadow, might feature a flower count of 1019. 10 chairs is an amount that one might imagine around a large living room table. And 83 cars might summon the mental picture of a large parking lot. The amounts were presented in a non-linear sequence, to make it harder for participants to extrapolate their previous answer to the next item.

The instructions continued with: <Name> says: "<Quantifier> <category> are <color>." The quantifier was 'A few', 'Some' or 'Most'. In the first item, it read: Jan says: "<Quantifier> flowers are red.", in the second item: 'Mieke says: "<Quantifier> chairs are brown.", and in the third: Ingrid says: "<Quantifier> cars are green." Following the statement, the items featured the same three questions:

a. If this utterance of <Name> is appropriate, how many <color> <category> are there minimally at that location?
b. If this utterance of <Name> is appropriate, how many <color> <category> are there maximally at that location?
c. If this utterance of <Name> is appropriate, what is the most likely amount of <color> <category> at that location?'

The participants filled in the questionnaire with the three items, covering the three amounts and categories (i.e. resp. 1019 flowers, 10 chairs and 83 cars). Each item of a questionnaire regarded the same quantifier, resulting in three between-subjects conditions: A Few, Some and Most. The conditions only differed in which quantifier was presented in the statements. For Condition Some, for example, the statements read:

(1) Jan says: "Some flowers are red."
(2) Mieke says: "Some chairs are brown."
(3) Ingrid says: "Some cars are green."

Results and Discussion

Four participants did not answer every question (with a numeric amount), one participant answered every question with the population size and 22 participants reported a most likely value outside their reported [min;max] zone. Therefore they were excluded from further analyses. Two participants responded on certain questions with two adjacent amounts (e.g., '7 or 8'), for those answers we used the average of the two amounts (i.c., 7.5). All of the remaining 189 participants (i.e. 62 in Condition A Few, 66 in Some and 61 in Most) reported minimal values lower than the reported maximal values.

The Most condition was included as a test of the protocol we used. The semantic meaning of most is captured in its name, and in this protocol indicates 'more than half'. 154 out of 183 (84%) minimal values for most were indeed higher than half of the population size. The minimal values that were lower than expected might be explained by the comment of a few participants that the number of subgroups in the population is unknown. They seem to have interpreted most as indicating 'the largest of all subgroups' (e.g., "Most flowers are red." interpreted as: "There are more red than any other color flowers."), which might have a size lower than half of the total population if there are more than two subgroups. This alternative interpretation does not interfere with our current basic investigation of the numerical description of some and a few, yet it could be subject to future, more in-depth research.

Table 1: The mean minimum (Min), most likely (ML) and maximum (Max) value for the three quantifiers for each of the three population categories (scaled to 100).

	Value	A Few	Some	Most
10 chairs	Min	22	22	59
	ML	40	41	72
	Max	64	68	89
83 cars	Min	6	10	55
	ML	27	32	75
	Max	58	66	96
1019 flowers	Min	4	10	56
	ML	29	36	78
	Max	59	72	98

Table 1 summarizes the different conditions. All numbers are scaled to 100, in order to make comparisons between the different population values easier. In this view on the mean numbers, some seems to be described with higher values than a few and most features higher mean values than the other two quantifiers.

We tested the differences between the three conditions with the non-parametric Mann-Whitney U on the data scaled to 100. In looking at all population categories together, some is indeed seen as significantly higher than a few: in its minimal value ($U = 16169.50$, $p = .02$), in its most likely value ($U = 15791.50$, $p = .008$) and its maximum ($U = 15430.00$, $p = .003$). The values of most are significantly higher than those of a few (resp. U = 776.50, 758.00 and 6638.00, all with $p < .001$) and of some (resp. U = 1199.00, 685.5 and 7944.50, all with $p < .001$). The range (i.e. maximum minus minimum) of some is also broader than that of a few (16426,00, $p = .03$), and the range of most is broader than that of both a few ($U = 14588.50$, $p < .001$) and some ($U = 12738.50$, $p < .001$).

At the population level the results tell a more nuanced story (see Table 2). In each case some still produces higher values than a few, yet only in three cases these differences remain significant: the most likely value in the population category of 83 cars, and the most likely value and maximum in 1019 flowers. The difference in maximal value borders significance in the population category of 83 cars, as well as the minimal value and the range in 1019 flowers. The differences in numerical interpretation between some and a few appear to be bound by certain contextual factors such as population size. Not only do discrepancies in minimum, most likely value, maximum and range diminish when looking at a specific conceptual (e.g., ‘cars’) and/or numerical population size (e.g., 83). The number of these aspects that are found to be significantly different, grows with the population size.

Table 2: Mann-Whitney U comparison per population size between Some and A Few (from the data scaled to 100).

	Value	Min	ML	Max	Range
10 chairs	U	1933,50	1872,50	1806,50	1824,00
	P	.30	.21	.13	.14
83 cars	U	1910,00	1602,00	1736,00	1882,00
	P	.26	.02	.07	.26
1019 flowers	U	1731,00	1588,00	1572,50	1731,00
	P	.07	.01	.01	.07

At the population level the results tell a more nuanced story (cf. Table 2). In each case some still produces higher values than a few, yet only in three cases these differences remain significant: the most likely value in the population category of 83 cars, and the most likely value and maximum in 1019 flowers. The difference in maximal value borders significance in the population category of 83 cars, as well as the minimal value and the range in 1019 flowers. The differences in numerical interpretation between some and a few appear to be bound by certain contextual factors such as population size. Not only do discrepancies in minimum, most likely value, maximum and range diminish when looking at a specific conceptual (e.g., ‘cars’) and/or numerical population size (e.g., 83). The number of these aspects that are found to be significantly different, grows with the population size.

We hypothesized that some is more partitive, more scaled to the population size than a few. In this case, some should exhibit lower differences between population sizes (i.e. in the data scaled to 100). Within subjects, per description category (e.g., the minimum) we computed a variable D; e.g., $D_{min} = (cars_{min} – chairs_{min})^2 + (flowers_{min} – cars_{min})^2 + (flowers_{min} – chairs_{min})^2$. A lower value for D means that the data is more scaled, i.e. that the numerical description of the quantifier in question is more influenced by population size. The results show that some is significantly more scaled than a few in its minimum ($U = 1668.00$, $p = .04$) and its most likely value ($U = 1545.50$, $p = .009$). Its maximum is also more scaled, yet to a degree that only borders significance ($U = 1743.00$, $p = .07$). Overall we can conclude that the numerical description of some is more dependent on population size than that of a few.

Conclusions

The current study looks into a numerical description for several quantifiers on the none/some/all Horn scale, to form a basis for Dutch research on scalar implicatures. We enquired about the minimal, most likely and maximal value of quantifiers (most,) some and a few, given a certain population size (i.e. all). A more fine-grained methodology (e.g., not only focusing on one population size) is important given the diversity of findings in the literature. This approach paid off. Although the general picture is more or less straightforward, i.e. the parameter values for some are indeed higher than those for a few, our approach also showed some important nuances. This general trend was not significant for all population sizes. It seems that especially with lower population sizes, differences between a few and some are less pronounced. Analogous to the partitive attribute of the French certains (Pouscoulous et al., 2007), some appears to be more scaled than a few, its numerical description more influenced by population size. Yet since we focused on numerical descriptors, additional research is indicated to provide further evidence for the semantic implications of the current findings. For instance, there is clearly a generalizability issue. The study only focuses on three distinct cognitive categories and one can wonder to what extent our findings are generalizable when other categories or even other representatives of the same categories would have been used.

Acknowledgments

This research was carried out with the financial support of the National Council for Scientific Research – Flanders, Belgium (FWO grant G.0634.09).

References

Ameel, E., Storms, G., Malt, B., & Sloman, S. (2004). Linguistic diversity and the bilingual lexicon: The Belgian story. *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 49-54).

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. *Journal of Experimental Psychology*, *102*, 335-338.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*, 434-463.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The quarterly journal of experimental psychology*, *61*, 1741-1760.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology, 54*, 128-133.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology*, *64*, 2352-2367.

Geurts, B. (2010). *Presuppositions and pronouns.* Oxford: Elsevier.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures. *Semantics and pragmatics, 2*, 1-34.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, I. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes*, *20*, 667-696.

Horn, L. R. (1972). *On the semantic properties of logical operators in English.* Doctoral dissertation, UCLA. Distributed by Indiana University Linguistics Club.

Janssens, L., & Schaeken, W. (2012). The role of task characteristics in children's scalar implicature production. *Proceedings of the Annual Conference of the Cognitive Science Society 2012* (pp. 509-514).

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81.

Levelt, W. J. M. (1989). Speaking: from Intention to Articulation. Cambridge, MA: MIT Press

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversation implicature.* Cambridge, MA: MIT Press.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua, 133*, 152-163.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language, 85*, 203-210.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*, 253-282.

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of autism and developmental disorders, 39*, 607-618.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*, 347-376.

Röhrig, S. (2010). *The Acquisition of Scalar Implicatures.* Göttinger Schriften zur Englischen Philologie - Band 3. Göttingen: Universitätsverlag.

Van Dale (2014, April 13). Retrieved from http://www.vandale.be/opzoeken

Zevakhina, N. (2012). Strength and similarity of scalar alternatives. In A. Aguilar Guevara, A. Chernilovskaya & R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16: Volume 2*, (pp. 647–658). MIT Working Papers in Linguistics.