Some implicatures reveal semantic differences

Arina Banga[1], Ingeborg Heutinck[2], Sanne M. Berends[2] & Petra Hendriks[2]

[1]*Radboud University Nijmegen* / [2]*University of Groningen*

1.   Introduction

In everyday conversation, sentence meanings are not always expressed explicitly, but can also be merely implied. For example, consider the following sentence:

(1)   Some elephants have trunks.

Taken literally, this sentence is true because in the present world elephants generally have trunks. From this it logically follows that at least some elephants have trunks. However, many people would consider this sentence false. For them, it would be more appropriate to use the quantifier *all* instead of *some*. According to standard pragmatic views, if the speaker uses the underinformative or weak form on a scale of informativeness, the listener may assume that the speaker is not in a position to use the informative or strong form on the same scale (Grice 1975). Thus, using the weak form *some* indicates that the strong form *all* is not applicable here. This leads a listener to interpret (1) as meaning that not all elephants have trunks. This inference is called a scalar implicature. Because the implicature is not true in the present world, many people judge (1) to be false. Scalar implicatures can be derived on the basis of a variety of scales. In (1), the relevant scale of informativeness is <*some*, *all*>. Other scales are <*or*, *and*>, <*sometimes*, *always*>, and <*may*, *must*>. A characteristic of these scales is that their elements are ordered by entailment relations. Elements on the right side of the scale are informationally stronger, and therefore logically entail the weaker ones.

Recently, several studies have experimentally investigated the ability of adults and children to compute scalar implicatures. These studies reveal that children compute fewer implicatures than adults do (e.g., Noveck 2001), and hence are said to be more logical than adults, although even adults do not compute implicatures in all cases. Furthermore, it is shown that children's performance improves when the experimental conditions are manipulated (Guasti et al. 2005; Papafragou & Musolino 2003; Pouscoulous et al. 2007), suggesting that cognitive resources may be a factor in children's limited ability to compute scalar implicatures. In line with this view, it has been argued that the observed variation among adults is caused by differences in working memory capacity as well (Feeney et al. 2004).

In this article, we discuss two experiments investigating the computation of scalar implicatures by Dutch adults. The aim of the experiments is to shed more light on the observed variation among individuals and across studies.

2.   Experiment 1

2.1. *Introduction*

If limited working memory capacity is responsible for children's generally lower rate of implicatures as well as for the observed variation among adults, it is expected that working memory capacity correlates with rate of implicatures in adults. We tested this in an experiment with Dutch adults.

2.2. *Method*

2.2.1.   *Participants*

Seventy-three adult native speakers of Dutch (age range 18-29; mean age 23;7) participated in this study: 19 with low working memory capacity, 39 with middle working memory capacity and 15 with high working memory capacity. Working memory capacity was assessed before the actual experiment took place (see below). Most participants were university students, who participated voluntary. None of the participants had (a history of) neurological, auditive or visual problems.

2.2.2.   *Materials and procedure*

To investigate the verbal working memory capacities of the participants, we tested participants on the reading span task of Van den Noort et al. (2008), which is a Dutch version of Daneman and Carpenter (1980). This task required

participants to read aloud series of sentences, increasing in length from 2 to 6 sentences, while trying to remember the last word of each sentence. At the end of the series of sentences, participants had to repeat back these words in the correct order. Participants were categorized as high-, middle- or low-spanners according to the criteria in Van den Noort et al. After the working memory test, the middle-spanners (but not the participants who were categorized as low-spanners or high-spanners) did a pre-test task (a verbal reasoning task) which is not relevant for the present discussion.

Following the working memory task (for all participants) and the pre-test task (for middle-spanners only), all participants took part in the actual experiment, an auditory Truth-Value Judgment Task. This task required participants to judge whether a given statement is true or false. Three experimental factors were included, namely Situation Type (an 'all' situation that makes the proposition true in all cases, versus a 'subset' situation that makes the proposition true in only a subset of the cases), Scale (determiner scale <*sommige*, *alle*> versus adverb scale <*soms*, *altijd*>) and Strength (strong versus weak term on a scale). These factors resulted in the following types of statements (in their English translations):

	'All' situation		*'Subset' situation*
(2)	All fridges have doors.	(6)	All children are blonde.
(3)	Some elephants have trunks.	(7)	Some books have pictures.
(4)	Airplanes always have wings.	(8)	Flowers always are yellow.
(5)	Giraffes sometimes have long necks.	(9)	Birds sometimes live in cages.

The statements in (3) and (5) are underinformative, as they are true in all situations but contain the weak term on the scale. Hence, they may be considered infelicitous. The response on these statements depends on whether the participants compute an implicature or not. If participants do not compute an implicature, they are expected to judge these statements to be true. If they compute an implicature, they are expected to judge these statements to be false.

Participants received 24 test sentences, which were adapted from Feeney et al. (2004), whose test sentences are a subset of Noveck's (2001) materials. For each of the 8 conditions above, 2 statements were included. In addition, 8 absurd statements were included, balancing 'true' versus 'false' responses (2 for each scalar term, e.g., *Some birds have telephones*). Furthermore, participants received 12 filler items containing a negation (6 true, 6 false): 6 of them contained the sentential negation *geen* ('no'), and 6 contained the sentential negation *nooit* ('never'). In total, therefore, each participant received 36 statements.

Each test sentence was presented with both *sommige* ('some') and *alle* ('all'), although each participant received only one of the two items. The same holds for the test sentences with *soms* ('sometimes') and *altijd* ('always'). The same 12 filler items were presented to each participant. This resulted in two versions of the materials. For each of these versions two random orders were prepared to control for order effects, resulting in a total of four lists. The test sentences were pre-recorded and were produced by a woman with a neutral accent who did not have any knowledge of the goal of the experiment. Sentences were presented at a regular speed, with an interval of three seconds separating the sentences. Participants had to say "true" or "false" to the statement, after which the experimenter wrote down their response. Each test session started with two practice items (one true, one false) to familiarize participants with the experiment.

2.3. *Results*

The mean percentages of logical responses on underinformative statements with *sommige* and *soms* (cf. (3) and (5)) versus informative statements with these terms (cf. (7) and (9)) are graphically presented in Figure 1.
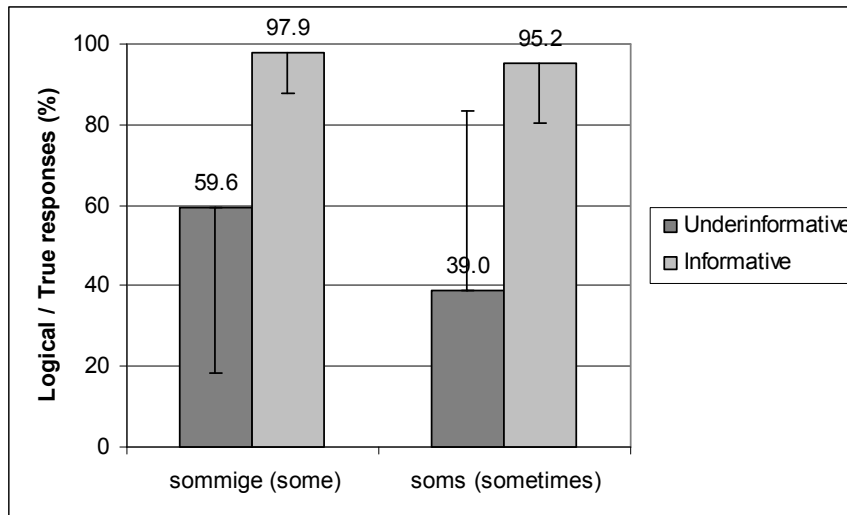

Figure 1: Percentages of logical responses for the scalar terms 'sommige' and 'soms' in underinformative versus informative statements. The error bars indicate one standard deviation from the average.

General Linear Model Repeated Measures ANOVA test statistics were run on these logical response percentages with Situation Type ('all' situation versus 'subset' situation), Scale (determiner versus adverb), and Strength (strong versus weak) as within-participants factors and Span (low versus middle versus high) as a between-subjects factor. For the participants with middle verbal working memory span, Pre-test Task was a between-subjects factor as well.

The main effect of Scale is significant, $F(1,70) = 11.552$, $p = .001$. The determiner scale elicited more 'true' responses than the adverb scale. Furthermore, the main effect of Strength is significant, $F(1,70) = 78.965$, $p < .001$. The statements containing strong scalar terms elicited a higher number of 'true' responses than the statements containing weak scalar terms. The main effect of Situation Type is also significant, $F(1,70) = 134.852$, $p = <.001$. Statements about 'all' situations elicited more 'true' responses than statements about 'subset' situations. The interaction between Scale x Strength is significant, $F(1,70) = 6.751$, $p = .011$. The adverb scale elicited a higher number of 'true' responses than the determiner scale, except for statements about 'all' situations. For these statements, the determiner scale elicited more 'true' responses. The interaction between Scale x Situation Type is significant, $F(1,70) = 5.989$, $p = .017$. The adverb statements about 'all' situations elicited a lower number of 'true' responses than the other conditions. The interaction between Strength x Situation Type is significant, $F(1,70) = 75.902$, $p < .001$. The statements about 'all' situations that contain a weak determiner elicited a lower number of 'true' responses than the other conditions. The interaction between Scale x Strength x Situation Type is significant, $F(1,70) = 4.773$, $p = .032$. The statements about 'all' situations that contain the weak adverb elicited fewer 'true' responses than the other conditions. The interaction between Scale x Pre-test Task is not significant, $F(1,37)= 1.650$, $p = .207$, neither is the interaction between Strength x Pre-test Task, $F(1,37)= 0.478$, $p = .494$, nor is the interaction between Situation Type x Pre-test Task, $F(1,37)= 0.855$, $p = .361$. The interaction between Scale x Strength x Pre-test Task is not significant, $F(1,37)= 0.819$, $p = .371$, neither is the interaction between Scale x Situation Type x Pre-test Task , $F(1,37)= 1.121$, $p = .297$, nor is the interaction between Strength x Situation Type x Pre-test Task, $F(1,37)= 2.026$, $p = .163$. The interaction between Pre-test Task x Scale x Strength x Situation Type is not significant, $F(1,37)= 0.017$, $p = .896$. Because no effect of the factor Pre-test Task was found, the middle-spanners can be compared with the low-spanners and the high-spanners, who did not receive a pre-test task. The interaction between Scale x Span is not significant, $F(2,70) = 2.063$, $p = .135$, neither is the interaction between Strength x Span, $F(2,70) = 0.485$, $p = .618$, nor is the interaction between Situation Type x Span, $F(2,70) = 1.552$, $p = .219$. The interaction between Scale x Strength x Span is not

significant, $F(2,70) = 1.995$, $p = .144$, neither is the interaction between Scale x Situation Type x Span, $F(2,70) = 3.026$, $p = .055$, nor is the interaction between Strength x Situation Type x Span, $F(2,70) = 0.189$, $p = .828$. The interaction between Scale x Strength x Situation Type x Span is not significant, $F(2,70) = 0.586$, $p = 0.559$.

2.4. *Discussion*

Informative statements with *sommige* ('some') and *soms* ('sometimes') were judged to be true in almost 100% of the cases (see Figure 1), whereas underinformative statements with these scalar terms were judged to be true significantly less often. The rejections of underinformative statements indicate that adult speakers of Dutch compute implicatures with *sommige* and *soms*, and frequently interpret these items as meaning 'some but not all' and 'sometimes but not always', respectively.

No significant effects of working memory capacity were found. With *sommige* and *soms*, low-spanners gave pragmatic (implicature) responses only (n=6 for *sommige*/n=11 for *soms*), mixed responses (n=5/n=4) as well as logical responses only (n=8/n=4). The same variation was observed among middle-spanners (pragmatic responses only: n=8/n=20; mixed responses: n=12/n=8; logical responses only: n=20/n=12) and high-spanners (pragmatic responses only: n=6/n=7; mixed responses: n=4/n=2; logical responses only: n=5/n=6). So even though adults did not compute implicatures in all cases, adults with a low working memory span did not compute implicatures at a significantly different rate than adults with a high working memory span. Thus, our results disconfirm the conclusions of Feeney et al. (2004) that working memory capacity has an effect on implicature computation.

On average, participants gave logical responses with *sommige* in 58.8% of the cases and with *soms* in 39.2% of the cases, corresponding to a rate of implicatures of 41.2% with *sommige* and 60.8% with *soms*. This difference is significant, and suggests that particular properties of the scalar term or the sentence containing the scalar term may be responsible for the different rate of implicatures. One possibility is that the difference between responses on statements with *sommige* and *soms* does not really reflect a difference in implicature computation, but rather is caused by a generic interpretation of the statements with *soms*. If a sentence such as *Giraffes sometimes have long necks* is assigned a generic interpretation (meaning that giraffes have long necks on certain days but have short necks on other days), participants are expected to answer 'false' too, just as when they compute an implicature. This will result in a higher rate of 'false' responses with *soms* than with *sommige*. An alternative

explanation for the higher rate of 'true' responses with *sommige* than with *soms* is that this is caused by particular properties of the scalar term *sommige*. In a second experiment, we investigate this latter possibility.

3.   Experiment 2

3.1. *Introduction*

Might the rate of implicatures be dependent on the scalar term involved? In an experiment with French children and adults, Pouscoulous et al. (2007) investigated the availability of implicatures with the French scalar terms *certains* and *quelques*, both meaning 'some'. Pouscoulous et al. assume that *quelques* is easier to process than *certains*. As evidence for this assumption, they point at a series of observations showing that *certains* is less frequent than *quelques* in children's written production, children's books and adult language, and that young children understand *quelques* more readily than *certains*. Indeed, they find that children, but not adults, compute more implicatures with *quelques* than with *certains*. Pouscoulous et al. attribute this difference in implicature computation to the lexical complexity of *certains*. Because *certains* receives a partitive interpretation, it is harder to interpret, they claim. As a consequence, children have fewer resources left to compute the implicature. In contrast, the adults in their experiment "have enough resources available to appear unaffected by our choice of word" (p. 370).

Like French, Dutch also has several terms to express existential quantification. It is not unusual to translate English *some* with Dutch *sommige*. However, Dutch also has the form *enkele*. According to De Jong (1983), *sommige* gives rise to a presupposition (*sommige wel* 'some do' implies *andere niet* 'others don't'), whereas *enkele* can also be used non-presuppositionally. The presupposition of *sommige* follows from its partitive interpretation ('a part of the'), she claims. As such, *sommige* is equivalent to the partitive construction *enkele van de* ('some of the'). This explains why *sommige* and *enkele van de* cannot appear in the following existential sentence, whereas *enkele* can:

(10)   Er liggen enkele/*sommige/*enkele van de bunkers in het duingebied.
       There lie some (of the) bunkers in the dune-area

De Hoop and Kas (1989; see also De Hoop 1992), on the other hand, argue that *sommige* differs from *enkele,* not because *sommige* is presuppositional, but because *sommige* is not a purely quantitative determiner. *Sommige* refers to a

certain subset of the set referred to by the noun. The meaning of *sommige* is such that the members of this subset must be related to each other by an implicit or explicit property. In the sentence *Sommige katten zijn doof* ('some cats are deaf'), the set of cats that are deaf must have some property in common, for example the property that they are white. Because of this non-quantitative aspect of its meaning, *sommige* cannot appear in existential sentences. The unacceptability of *enkele van de* in (10) has a different explanation, according to De Hoop and Kas: morphosyntactic partitivity implies the existence of a finite set, which is not given in (10).

Scalar and other conversational implicatures have been argued to be non-detachable: every expression with the same meaning carries the same implicature (Grice 1975:39). This property of implicatures can be exploited to investigate the existence of subtle meaning differences between lexical items. If De Jong is correct in her claim that *sommige* and *enkele van de* carry the same meaning, we expect to find the same rate of implicatures with *sommige* and *enkele van de*. Furthermore, these terms may give rise to a different rate of implicatures compared to *enkele*, leading to more implicatures for *sommige* and *enkele van de* if a partitive interpretation indeed raises the salience of the larger set and therefore makes it easier for adults to compute the implicature (cf. Pouscoulous et al. 2007:360). On the other hand, if De Hoop and Kas are right, we may see a different rate of 'true' responses with *sommige* compared to *enkele*, because *sommige* is not a purely quantitative determiner. De Hoop and Kas (1989:47) propose that the truth conditions of *sommige* include both quantitative and qualitative requirements. Because of the quantitative requirement that the subset referred to by *sommige* must have at least one member, *sommige* might be able to evoke a scale <*sommige, alle*> and hence give rise to an implicature. However, in addition to this quantitative requirement, *sommige* is also argued to be subject to the qualitative requirement that the members of the subset selected by *sommige* must have an implicit or explicit property in common. This qualitative requirement may have different effects: it is conceivable that participants reject underinformative statements with *sommige* like *Sommige boeken hebben bladzijden* ('some books have pages') because they fail to identify such a property. In this case, we expect responses resembling those associated with the computation of an implicature. Alternatively, participants may succeed in finding a common property defining a subset even for what we consider to be 'all' situations (e.g., the property of being made of paper rather than being digital, when talking about the subset of books that have pages). In that case, there is no need to compute an implicature and we expect a logical response. Consequently, we may see either fewer or more 'true' responses with *sommige* than with *enkele*. A further prediction is that under De Hoop and Kas' account we do not expect any differences in the rate of implicatures between partitive and non-partitive

constructions, because morphosyntactic partitivity does not affect the quantitative properties of scalar terms. We tested these predictions in an experiment similar to Experiment 1.

3.2. *Method*

3.2.1. *Participants*
Thirty-two adult native speakers of Dutch (age range 20-36, mean age 27;2) participated in this experiment. The participants in Experiment 2 were university students who had not participated in Experiment 1. None of them had (a history of) neurological, auditive or visual problems.

3.2.2. *Materials and procedure*
Experiment 2 consisted of a Truth-Value Judgment Task. No pre-test tasks were administered. The experiment included three factors, namely Situation Type ('all' situation versus 'subset' situation), Scalar Term (*sommige* ('some') versus *enkele* ('some')) and Partitivity (with versus without the partitive construction *van de* ('of the')), resulting in the following types of statements:

'All' situation
(11) Enkele olifanten hebben slurven. ('some elephants have trunks')
(12) Sommige boeken hebben bladzijden. ('some books have pages')
(13) Enkele van de giraffen hebben lange nekken. ('some of the giraffes have long necks')
(14) Sommige van de katten hebben oren. ('some of the cats have ears')

'Subset' situation
(15) Enkele boeken hebben plaatjes. ('some books have pictures')
(16) Sommige vogels leven in kooien. ('some birds live in cages')
(17) Enkele van de bloemen zijn geel. ('some of the flowers are yellow')
(18) Sommige van de jurken hebben ritsen. ('some of the dresses have zips')

Participants received a total of 48 statements, 16 of which were test items and 32 were filler items. The 16 test items consisted of 2 statements for each of the 8 conditions above. Of the 32 filler items, 16 contained the universal quantifier *alle* ('all') and 16 contained the negative determiner *geen enkele* ('not any'). No absurd statements were included. Four versions of the experiment were constructed by a Latin square design, so that each version contained two items of each condition and one version of each item. The fillers were the same for each version of the test. The procedure was the same as in Experiment 1.

3.3. *Results*

The mean percentages of logical responses with the four scalar terms are graphically presented in Figure 2.
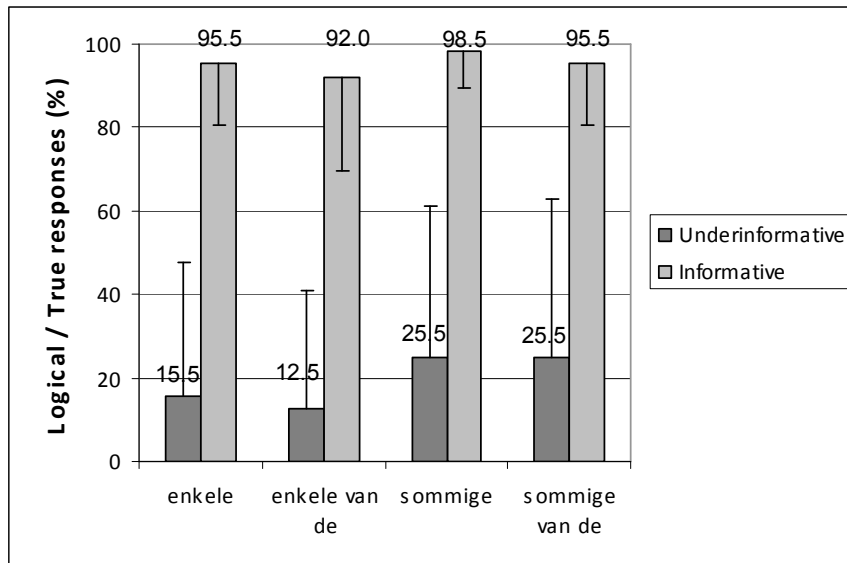

Figure 2: Percentages of logical responses for four scalar terms in Dutch meaning 'some', in underinformative versus informative statements. The error bars indicate one standard deviation from the average.

General Linear Model Repeated Measures ANOVA test statistics were run on the percentages of 'true' responses, with Situation Type ('all' versus 'subset' situation), Scalar Term (*sommige* ('some') versus *enkele* ('some')), and Partitivity (with versus without the partitive construction *van de* ('of the')) as within-participants factors. The main effect of Scalar Term is significant, $F(1,31) = 9.266$, $p = .005$. The scalar term *sommige* elicited more 'true' responses than the scalar term *enkele*. The main effect of Partitivity is not significant, $F(1,31) = 1.130$, $p = .296$. The main effect of Situation Type is significant, $F(1,31) = 246.559$, $p < .001$. Statements about 'all' situations elicited more 'true' responses than statements about 'subset' situations. The interaction between Scalar Term x Partitivity is not significant, $F(1,31) = 0.088$, $p = .768$, neither is the interaction between Scalar Term x Situation Type, $F(1,31) = 2.627$, $p = .115$, nor is the interaction between Partitivity x Situation Type, $F(1,31) = 0.018$, $p = .745$. The

three-way interaction between Scalar Term x Partitivity x Situation Type is also not significant, $F(1,31) = 0.081$, $p = .778$.

3.4. *Discussion*

Statements with *enkele* ('some') gave rise to significantly more 'false' responses (indicating an implicature) than statements with *sommige* ('some'). The presence or absence of a partitive construction, in contrast, had no effect on the rate of 'false' responses; *enkele van de/sommige van de* did not result in more or fewer 'false' responses than *enkele/sommige*.

These results are compatible with the predictions generated on the basis of the analysis of De Hoop and Kas (1989) of the semantic difference between *sommige* and *enkele*, but seem unexplainable from the analysis of De Jong (1983). Thus, the results of Experiment 2 provide independent evidence that *sommige (van de)* and *enkele (van de)* differ semantically. The lower rate of 'false' responses with *sommige* compared to *enkele* is explained by the fact that *sommige* is not a purely quantitative determiner but requires that the members of the set referred to share a particular property. From our experimental results, we are unable to determine why participants accepted or rejected statements with *sommige*. It is possible that for some of the underinformative statements with *sommige* participants were able to identify a relevant property, accepting these statements without computing an implicature. For the other underinformative statements, participants may have failed to identify a relevant property, rejecting these statements without computing an implicature. However, we cannot entirely rule out the possibility that participants simply ignored the qualitative requirement for *sommige* and did compute an implicature, although for some reason at a lower rate than with *enkele*.

If French *certains* resembles Dutch *sommige*, qualitative aspects of the meaning of *certains* may also yield an explanation for children's higher rate of errors with *certains* compared to *quelques* in the study of Pouscoulous et al. (2007). Note that Pouscoulous et al.'s explanation in terms of partitivity (which is similar to De Jong's account) would not be able to explain our results.

The different behaviour of *sommige* ('some') and *soms* ('sometimes') in Experiment 1 is also explainable from the partly qualitative nature of *sommige*. Because *soms* does not require that the members of the subset it selects are related to each other by a particular property, *soms* is expected to behave more like *enkele* than like *sommige*. Indeed, *soms* in Experiment 1 gave rise to more 'false' responses than *sommige*, analogous to *enkele* in Experiment 2.

In Experiment 2, *enkele* and *sommige* in partitive constructions led to 'false' (implicature-like) responses only (n=26 for *enkele*/n=21 for *sommige*), mixed responses (n=4/n=6) as well as 'true' (logical) responses only (n=2/n=5). *Enkele*

and *sommige* in non-partitive constructions also led to 'false' responses only (n=25/n=20), mixed responses (n=4/n=8) as well as 'true' responses only (n=3/n=4). So in Experiment 1 as well as in Experiment 2, individual variation was observed both between and within participants.

Although Experiment 1 and 2 were highly similar, participants gave 'true' answers with *sommige* in 58.8% of the cases in Experiment 1, but in only 25.0% of the cases in Experiment 2. Apart from the different scalar terms used and the presence of pre-test tasks in Experiment 1, the main difference between the two experiments lies in the filler items used. Using absurd filler items (as we did in Experiment 1, following Feeney et al. 2004, and Noveck 2001) may have caused the participants to be more tolerant towards underinformative statements with *sommige*, thus suspending the need to either identify the relevant property or compute an implicature. Pouscoulous et al. (2007) already emphasized the importance of the presence or absence of filler (or, in their terminology, distractor) items for children's computation of implicatures. Our results with adults suggest that also the nature of the filler items is important.

4.   Conclusion

We carried out two experiments investigating the influence of working memory capacity and semantic properties of the scalar terms on the computation of scalar implicatures by Dutch adults. In Experiment 1, we found substantial individual variation among adult participants: some participants computed implicatures in all cases, whereas others gave mixed responses or logical responses only. However, we did not find any evidence for the influence of working memory capacity on the rate of implicatures. We did find an effect of scalar term, with *soms* ('sometimes') giving rise to more implicature-like responses than *sommige* ('some'). This result led us to perform a second experiment. In Experiment 2, we found that the purely quantitative determiner *enkele* ('some') gave rise to more implicature-like responses than the partly qualitative determiner *sommige* ('some'), suggesting that semantic differences between scalar terms affect the computation of implicatures in adults. Morphosyntactic partitivity, on the other hand, did not affect the computation of implicatures. Thus, subtle semantic differences between scalar terms may be responsible for part of the observed variation across studies. However, the particular choice of filler items may have an even larger effect. This latter conclusion is supported by the highly different results we found for the same test items involving *sommige* in Experiment 1 and 2.

Acknowledgments

References

Daneman, M. & P. A. Carpenter. 1980. "Individual differences in working memory and reading". *Journal of Verbal Learning and Verbal Behavior* 19. 450-466.

Feeney, A., S. Scrafton, A. Duckworth, & S. J. Handley. 2004. "The story of some: Everyday pragmatic inference by children and adults". *Canadian Journal of Experimental Psychology* 58:2. 121-132.

Grice, H. P. 1975. "Logic and conversation". *Syntax and semantics: Speech acts. Volume 3* ed. by P. Cole & J. L. Morgan, 41-58. New York: Academic Press.

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

de Hoop, H., & M. Kas. 1989. "Sommige betekenisaspecten van enkele kwantoren, oftewel: enkele betekenisaspecten van sommige kwantoren". *TTT* 9:1. 31-49.

de Hoop, H. 1992. *Case configuration and noun phrase interpretation*. PhD diss., University of Groningen.

Jong, F. de. 1983. "Sommige niet, andere wel: de verklaring van een raadselachtig verschil". *Glot* 6. 229-246.

Noort, M. W. M. L. van den, M. P. C. Bosch, M. Haverkort, & K. Hugdahl. 2008. "A standard computerized version of the reading span test in different languages". *European Journal of Psychological Assessment* 24:1. 35-42.

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

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

Pouscoulous, N., I. A. Noveck, G. Politzer & A. Bastide. 2007. "A developmental investigation of processing costs in implicature production". *Language Acquisition* 14:4. 347-375.