In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds., 2014). Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Electrophysiology of Pragmatic Processing: Exploring the Processing Cost of the Scalar Implicature in the Truth-Value Judgment Task

Maria Spychalska* (maria.spychalska@rub.de), Jarmo Kontinen* (jarmo.kontinen@rub.de), Markus Werning* (markus.werning@rub.de)
*Institute of Philosophy II, Ruhr University Bochum

Abstract

Most theoretical as well as empirical work regarding the scalar implicature *not all* of the quantifier *some* has focused on the controversy of whether this implicature is generated by default or based on context. Independently of this question, it can be also asked whether this scalar implicature contributes to the truth-conditional content of sentences. We present results of an ERP study which tackles both these questions. We adopt a sentence-picture verification paradigm to investigate whether the violation of this implicature, when it is based on short-term memory rather than semantic memory, elicits N400 or late positivity (e.g. P600) effects. Additionally, we test if the truth-value judgments of the pragmatically infelicitous sentences correlate with the elicited ERP components.

Keywords: scalar implicature; truth-value judgment; N400 effect; Post-N400-Positivity; P600

Introduction

In this paper we present results of the EEG experiment in which we investigate how people process the scalar implicature *nicht alle* (*not all*) of the German quantifier *einige* (*some*) in a truth-value judgement task. Scalar implicatures are pragmatic inferences, i.e. they are based not only on the literal meaning of a given linguistic expression, but also on the additional assumption that the speaker is contributing to a conversation in a cooperative way, in particular by providing an appropriate amount of information. In standard semantics *Some As are B* is understood to be equivalent to *There are As that are B* and, hence, is semantically compatible with *All As are B*. This standard meaning of *some* is usually referred to as its *logical (semantic) meaning*. However, if a speaker says *Some of the students passed the exam*, the addressee is in a position to infer that *Not all students passed the exam*, since, according to the cooperativity requirement, the speaker is expected to either utter a more informative statement with the quantifier *all* if she knew that it was true, or reveal her lack of knowledge on this matter. The meaning of *some*, when enriched with the implicature *not all*, is often referred to as its *pragmatic meaning*.

One particularly often addressed question regarding scalar implicatures is the *default* vs. *context-based* controversy. According to the default theory the implicature *not all* is generated by the lexical item *some* by default, i.e. locally and more or less automatically, albeit it may be canceled in special circumstances (Horn, 1984; Levinson, 1983; Chierchia et al., 2012). The proponents of the second approach (Bott & Noveck, 2004; Breheny et al., 2006) postulate that the logical reading is the default interpretation of *some*, whereas the scalar implicature result from complex and global reasoning processes that are based on context.

Somewhat orthogonal to this debate, the emphasis in the philosophy of language concerning the semantics-pragmatics distinction has recently shifted towards the role pragmatic processes play in establishing *truth-conditions* of sentences. From this perspective, one can ask whether scalar implicatures are purely post-semantic processes that do not contribute to *what is (literally) said*, or whether they enter the content of *what is said* and hence influence the truth-conditions. Minimal semanticists (Cappelen & Lepore, 2005; Borg, 2012) argue that a sentence's truth-conditions are determined solely by its compositional semantics, which is a function only of the semantic values of the sentence's constituents and the way they are syntactically combined. This traditional picture has been recently challenged. For instance, Recanati's truth-conditional pragmatics (Recanati, 2010) allows top-down pragmatic enrichment of *what is said*, therefore questioning the classical semantics-pragmatics distinction. According to this approach, not only *primary* pragmatic processes such as saturation of indexicals, but also *free* pragmatic processes, such as implicatures, can influence *intuitive truth-conditions* of sentences.

Most empirical work regarding scalar implicatures focuses on the default- vs. context-based controversy. In order to shed some light on this debate some researchers have investigated the processing of the implicature on the neural level, with the use of EEG (electroencephalography). Noveck & Posada (2003); Nieuwland et al. (2010) investigated whether the violation of the scalar implicature triggers a so-called N400 effect – an amplitude difference between the event-related potentials (ERPs), i.e. the time-locked EEG signals, in two compared conditions recorded roughly 400*ms* after the onset of the stimulus. This effect is known to be linked to the recognition of a semantic incongruence in language (Kutas & Federmeier, 2000). In the case of pragmatic infelicity, which is due to implicature violation, an N400 effect could be interpreted as evidence of an early incremental integration of the implicature into the sentence meaning, and thus support the default theory. Nieuwland et al. (2010) compared ERPs elicited by pragmatically less felicitous sentence-final predicates in underinformative sentences with *some*, e.g. *Some people have lungs* (*All people have lungs* would be more informative in this case), with felicitous sentence-final predicates in informative sentences with *some*, e.g. *Some people have pets*. They showed that people differ in the way they process these sentences, depending on their pragmatic language ability, which was estimated based on their score in the *Autism Spectrum Quotient Questionnaire* (AQ) (Baron-

Cohen et al., 2001). Individuals with low AQ (high pragmatic skills) were more sensitive to implicature violation, whereas participants with high AQ (low pragmatic ability) were more sensitive to the local lexical-semantic relationships in sentences, that were stronger in underinformative than in informative sentences. This relationship, measured by a so-called *Latent Semantic Analysis* (LSA) (Landauer et al., 1998), generally negatively correlates with the size the N400 effect.

In contrast, Politzer-Ahles et al. (2012) applied a paradigm in which the felicity of the use of the quantifier *some* was evaluated with respect to picture-models, e.g. pictures in which all agents were engaged in the same activity (*all*-models) or pictures in which only some of the agents were engaged in one activity and the rest in another activity (*some*-models). Measuring the ERPs at the onset of the quantifier, they showed that those quantifiers that were used in a pragmatically inconsistent (*some* in the case of *all*-models), but not semantically inconsistent way (*all* in the case of *some*-models), were associated with a *sustained negativity* effect, i.e. a prolonged negativity starting ca. 300*ms* post-stimulus onset that has often been observed in response to ambiguous words (Van Berkum et al., 2007).

The reported studies, however, have not given a satisfying response to a question, whether the violation of the scalar implicature, when unaffected by lexical-semantic constraints, can trigger the N400 effect independently of the general pragmatic skills (AQ score) of the subjects. Furthermore, since we know that people are usually divided in the way they evaluate the underinformative sentences (Bott & Noveck, 2004), we can also ask to what extent those intuitive truth-value judgments would affect the elicited ERP components in the case of the implicature violation. Nieuwland et al. (2010) did not provide any differential analysis of this sort, whereas in the study by Noveck & Posada (2003) no modulation of the ERPs by the truth-value judgements was found. Still, we can expect such a modulation to take place, especially if the process of evaluating pragmatic felicity is dissociated from semantic memory. From a logical perspective semantic truth is a case of a semantic congruence, whereas falsity represents a semantic violation. Thus, sentences evaluated as false should be associated with larger N400 ERPs than sentences evaluated as true. It is then also expected that in the case of pragmatically infelicitous sentences the elicited N400 ERPs should depend on the truth-value evaluation given by the subjects.

To investigate these questions we examined ERP effects associated with a violation of the scalar implicature in a sentence-picture verification paradigm. The reason to adopt a sentence-picture verification paradigm was to dissociate the process of implicature calculation from world-knowledge-based or semantic memory-based sentence evaluation and record ERPs elicited by pragmatic violations that are based on short-term memory. The ERPs were measured on the onset of the critical sentence-final nouns, that determine semantic truth and pragmatic felicity of the quantified sentences. In this way we focused on the felicity of the use of this noun in given sentences with *some* or *all* in the context of visually presented models. Sentences with *all* were used to control across evaluation-conditions for the model type, especially the priming effect of the critical word (see details below). Truth-value judgements were gathered along with the EEG data in order to test whether the semantic evaluation of the underinformative sentences correlates with the elicited ERP components. Additionally, we screened our participants with three parts of the *Wechsler Intelligence test* (WAIS) to measure their logical fluid intelligence (the matrix reasoning test), linguistic intelligence (the vocabulary test), and working memory (the digit span memory test) and with the *Autism Spectrum Quotient Questionnaire*.

Experimental design

In our experiment, the target sentences with the quantifier *some* and the control sentences with the quantifier *all* were evaluated with respect to visual models, consisting of five pictures. Each sentence was of form (1), where X denotes the critical noun.

(1) *Einige/alle Bilder enthalten Xs.*
Some / all pictures contain Xs.

Models were presented before the critical noun X, but after the first part of a sentence containing the quantifier phrase. In each model two different categories of objects were presented: one occurring in each of the pictures, the other occurring only in two or three of the pictures. There were three evaluation-conditions for each of the two quantifiers which gives us a 2×3 design: For the quantifier *some* (*S*-conditions) these were: *true and felicitous* (ST), *true and infelicitous* (*SI*), and *false* (SF). For *all* (*A*-conditions) there was one *true* condition (AT), and two *false* conditions: when the critical noun denoted one of the object categories presented in the pictures (AF) and when it denoted an object category that was not displayed in the pictures ($AF2$). The conditions *SI* and *AT* corresponded to the case when X denoted the object category that was contained by each of the pictures, *ST* and *AF* corresponded to the case when X denoted the object category that was contained by only a subset of the pictures, finally *SF* and $AF2$ corresponded to the case when X denoted an object category that was not displayed in any of the pictures. Thus, true and false *A*-conditions were reversed relative to true and infelicitous *S*-conditions with respect to whether the object category denoted by the critical noun occurred in all or only in some of the pictures. Furthermore, conditions *AF2* and *AF* differed with respect to priming of the critical word, in the same way as conditions *SF* and *SI/ST*, however, unlike *SF* and *SI/ST*, they did not differ with respect to the semantic status of the sentence. The structure of an experimental trial is illustrated in Figure 1, whereas Table 1 presents the evaluation-conditions for each of the critical words in the given example. Subjects gave the truth-value judgements at the end of each trial, after the critical word disappeared. No feedback was given throughout the experiment and the subjects were asked to follow their intuition in the truth-value judgment task.

Figure 1: Time-course of an experimental trail

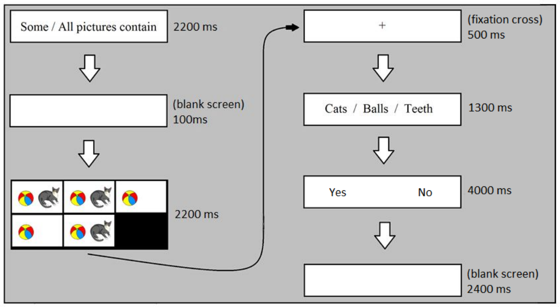

Table 1: Evaluation-conditions for both quantifiers with respect to the given picture and depending on the critical word

Quantifier	Balls	Cats	Teeth
Some	SI	ST	SF
All	AT	AF	AF2

Materials

For the preparation of the stimuli, we constructed eighty ordered triples of nouns $\langle n_1, n_2, n_3 \rangle$. All words were used in their plural form, were two-syllabic and had the length of $4-9$ characters; the compound nouns were excluded. The word frequency value[1] was kept between 8 and 17. In each triple the words were matched with respect to their length (maximal character difference was 4) and frequency (maximal value difference was 2). All nouns denoted concrete easily identifiable objects. For each triple we created individual pictures of n_1 and n_2 as single objects, and a picture of a pair of n_1 and n_2.

For each participant a different list of stimuli was generated from the base of noun-triples and pictures in a pseudorandom way. The same combination of nouns and pictures was used for each of the three conditions for a given quantifier. Depending on the evaluation-condition a different noun was displayed at the end of the trial. Thus, each picture set was seen three times, but each word only once. In each experimental trial two categories of objects were presented: one in each of the pictures, the other in three or two of the pictures. The position of these object categories (right- vs. left-hand side of the picture) was balanced across conditions but kept consistent within a trial. The proportion $5/3$ vs. $5/2$ was also balanced evenly per condition.

There were 40 trials per condition, which yields a total of 240 experimental trials, plus 60 filler trials, with quantifiers: *keine* (*no*), *die meisten* (*most*), *zwei* (*two*), *drei* (*three*), *vier* (*four*), *fünf* (*five*).

EEG recording and data processing

We measured fifty-seven (twenty-nine women) neurotypical, monolingual German native speakers (age: 18-44, mean: 24.2, SD: 4.4).[2] EEG was recorded from 64 active electrodes with a BrainAmp acticap EEG recording system. AFz was the ground electrode and FCz the physical reference. FT9, FT10, P09, PO10 were reprogrammed and used for controlling both vertical and horizontal eye-movements. The EEG was recorded with a sampling rate of 500 Hz and a bandpass filter of 0.53 (a time constant of $0.3s$ was used for a low cut-off) – $70Hz$. Impedance was kept below $5k\Omega$ for scalp electrodes and below $10k\Omega$ for EOG electrodes.

The EEG data were processed using Brain Vision Analyzer 2.0 software. We applied an off-line high cut-off filter at $40Hz$. Automatic raw data inspection rejected all trials with amplitude differences over $200\mu V$ in the interval of less than $200ms$, or with activity of less than $0.5\mu V$ in the interval of $100ms$. Maximal voltage step allowed was $50\mu V/ms$. Eye blinks were corrected using an independent component analysis. The data was off-line re-referenced to the linked mastoids comprising of TP9 and TP10. Segments from $200ms$ pre-target onset until $1000ms$ post-onset were averaged for every subject and every condition. Baseline correction used the $200ms$ interval preceding the onset of the stimulus. The minimal number of segments that was preserved in each condition was 26 out of 40 (60%).

Results

Behavioral results: "pragmatists" vs. "logicians" division

The analysis of the truth-value judgements revealed that our subjects were generally consistent in their choice of either the pragmatic or the logical interpretation of the quantifier *some*. Accordingly, we divided them into two groups based on their responses in condition *SI*. People who had at least 70% of pragmatic responses were called "pragmatists" ($N = 26$), whereas those who had at least 70% of logical responses were called "logicians" ($N = 28$). Applying the threshold of 70% resulted in an exhaustive division. The analysis of accuracy and reaction time is left out here due to the paper's space limits.

EEG results

For the statistical analysis of the EEG data we used a Matlab Fieldtrip package. We performed a non-parametric statistical procedure called cluster-based permutation test (Maris & Oostenveld, 2007) that allows to determine positive and negative clusters, i.e. collections of time-channel points where the measured amplitude in one condition is significantly higher respectively lower than in the compared condition.[3]

We pairwise compared averaged ERPs between *S*- and *A*-conditions separately, i.e. we compared *SI* vs. *ST*, *SF* vs. *ST* and *SF* vs. *SI* for the quantifier *some*, and *AF* vs. *AT*, *AF2* vs. *AT* and *AF2* vs. *AF* for the quantifier *all*. The crucial comparisons for testing our hypotheses were between conditions *SI* and *ST*, and between *SF* and *SI*. The comparison between conditions *false* and *true* for *some* (*SF* vs. *ST*) and between

[1] checked in the *Wortschatz Leipzig* corpus. The frequency *value* v of a word w is equal to the quotient of the log_2 of the frequency of the word "der" and the frequency of the word w in corpus

[2] Three people were excluded from the analysis due to technical problems.

[3] All the permutation tests were run on all channells and the whole epochs ($0-100ms$), with $\alpha = .025$, and 10000 permutations.

the conditions for the quantifier *all* served as control. After applying Bonferroni correction for three comparisons per quantifier, and for positive and negative clusters, our alpha level was $\alpha = .05/6 = .0083$. Our analysis revealed that for the quantifier *some* subjects' ERPs for critical words were significantly more negative in condition *infelicitous* (*SI*) than in condition *true* (*ST*) ($p < .001$), in the time window of 260-436*ms* post-onset, which corresponds to a standard N400 effect. This effect had a central topographical localization and was followed by a marginally significant centro-parietal positive cluster (500-624*ms*, $p < .0099$). Such late positivities have been often observed after N400 effects (see (Van Petten & Luka, 2012)), although their functional sensitivity is still unclear. Following Van Petten & Luka (2012) we adopt a theoretically neutral term *Post-N400-Positivity* (PNP) to refer to any positivity observed after an N400 effect.

The control comparison between conditions *false* (*SF*) and *true* (*ST*) for *some* predictably resulted in a significant N400 effect (236-492*ms*, $p < .0001$) followed by a significant PNP (520-998*ms*, $p < .0001$). More importantly, however, there was a significant N400 (254-514*ms*, $p < .0001$) and a significant PNP (544-964*ms*, $p < .0001$) in the second critical comparison, i.e. between conditions *false* and *infelicitous* for *some* (*SF* vs. *SI*). This means that even though the mean amplitude in condition *SI* was significantly larger in the N400 time-window than in *ST*, it was significantly smaller than in *SF*; however in the PNP time-window the mean amplitude was larger in *SI* than in *SF*, but smaller than in *ST*.

Similar effects were obtained for the quantifier *all*: Averaged ERPs in both *false* conditions were more negative than in the *true* condition in the time-window corresponding to the N400 effect: *AF* vs. *AT* (226-432*ms*, $p < .0001$) and *AF2* vs. *AT* (226-432*ms*, $p < .0001$). The N400 effects were followed by significant PNPs: *AF* vs. *AT* (440-960*ms*, $p < .0001$) and *AF2* vs. *AT* (510-998*ms*, $p < .0001$). The comparison between the two *false* conditions (*AF2* vs. *AF*) also resulted in a significant N400 effect (258-540*ms*, $p < .0001$) followed by a significant positivity effect (760-998*ms*, $p < .0036$). Note, that in this case, the positivity effect differed from the positivity effects obtained for the other comparisons, i.e. it started approximately 200*ms* later and had a more left-hemispheric extension.

The ERP effects depend on the truth-value judgement: the between-group differences

One of the main hypothesis of this study was that the ERPs elicited in the *infelicitous* condition should be modulated by subjects' truth-value evaluation of the underinformative sentences. Thus, we expected that "pragmatists" should get larger N400/P600 *SI* vs. *ST* effects than "logicians". The inspection of grand averages for both groups (Figure 2) suggested that this hypothesis could be sustained, and "pragmatists" and "logicians" indeed obtained different effects. The follow-up analysis involving cluster-based permutation tests that were performed separately for each of the two groups confirmed this hypothesis (Table 2)[4]. "Pragmatists" got a large *SI* vs. *ST* N400 effect lasting for almost 200*ms* with a global topographical extension, followed by an over 400*ms*-long wide-ranging PNP that was maximal at parietal regions, however, there were no significant effects in this comparison found for "logicians". In contrast, whereas "logicians" got a significant N400 effect and PNP in the comparison between conditions *SF* and *SI*, in the case of "pragmatists" only the N400 effect was significant. The N400 effect in this comparison obtained for "pragmatists" could be associated with the priming difference between conditions *SF* and *SI*.

To compare the size of the N400 effect and the PNP between the two groups we conducted independent t-tests. The two groups were compared with respect to their average amplitude differences between conditions *SI* and *ST*, as well as between *SF* and *SI*. The averages were calculated over all the electrodes involved in the effects, i.e. all electrodes that had at least one data point in the significant cluster.[5] The time-windows for computing the averaged size of the N400 and PNP effects were selected to be the time-windows of the corresponding significant clusters that were found by the permutation tests performed for all subjects, i.e. both groups jointly. According to our predictions, "pragmatists" got a significantly larger than "logicians" the *SI* vs. *ST* N400 effect ($t(52) = 5.392$, $p < .001$) as well as the *SI* vs. *ST* PNP effect ($t(52) = -3.506$, $p = .001$). There was also a significant correlation between the behavioral responses and these effects: The more often the subjects responded pragmatically, the larger was their N400 effect ($r = -.592$, $p < .001$), and PNP effect ($r = .467$, $p < .001$). The two groups also differed significantly with respect to the effects obtained for the comparison *SF* vs. *SI*. Both effects were significantly larger for "logicians" than for "pragmatists": the *SF* vs. *SI* N400 effect ($t(52) = -4.325$, $p < .001$) and the *SF* vs. *SI* PNP ($t(52) = 2.238$, $p = .03$). Both effects were also correlated with the behavioral responses: The more often the subjects responded pragmatically, the smaller was their PNP ($r = -.284$, $p = .037$) and the N400 effect ($r = .505$, $p < .001$). There were no between-group differences regarding the sizes of the N400 effects or PNPs obtained in the control comparison *SF* vs. *ST*, as well as in any of the comparisons between the *A*-conditions, except for the comparison *AF2* vs. *AF*.[6]

Finally, "pragmatists" and "logicians" did not differ in any of the personal characteristics measured: their intelligence, working memory or AQ. None of the measured variables was correlated with any of the ERP effects.

[4]The level of significance was additionally corrected for the two groups (per quantifier), i.e. $\alpha = .0083/2 = .00415$

[5]In each case we excluded the EOG electrodes, linked mastoids and, due to excessive artifacts, the anterior electrodes: AF7, AF8, Fp1, Fp2. For both N400 effects all remaining channels were included, except for F7. For the *SI* vs. *ST* PNP the *non*-involved electrodes were: F8, FC6, T8, F5, FT7, FT8; whereas for *SF* vs. *SI* PNP the *non*-involved electrodes were: F3, F8, AF3, AF4, F1.

[6]Here the late positivity cluster was significant for "logicians", but a similar cluster found by the permutation test for "pragmatists" was not significant.

Figure 2: Topographical maps of the effects in all S-comparisons plus the grand averages for all three conditions for the quantifier *some* at the electrode Cz

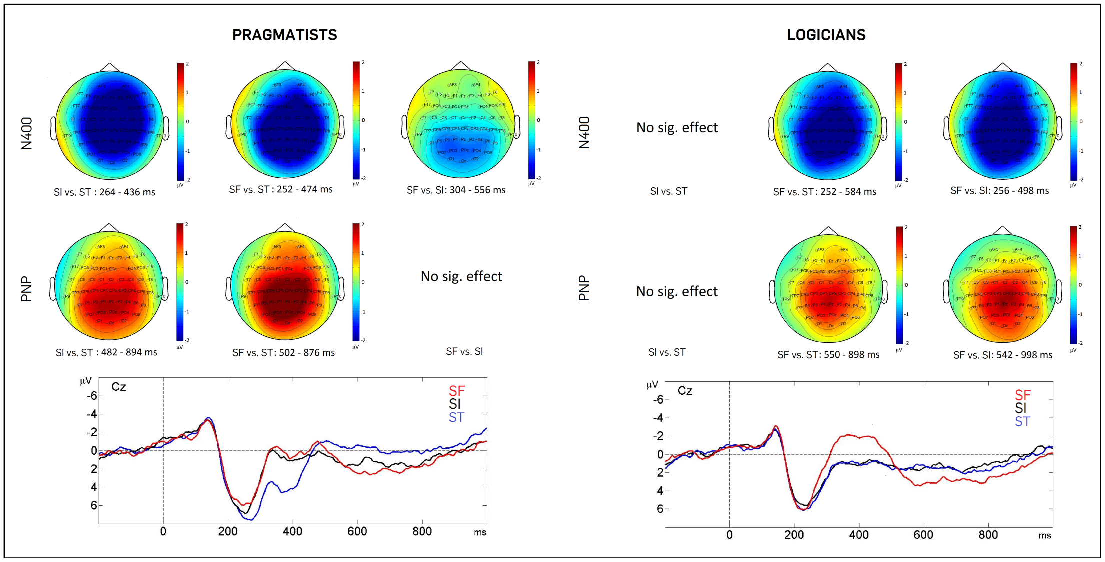

Table 2: Positive and negative clusters for each comparison for "pragmatists" and "logicians"

			SI vs ST	*SF vs ST*	*SF vs SI*	*AF vs AT*	*AF2 vs AT*	*AF2 vs AF*
Pragmatists	Negative clusters	time (ms)	264 – 436	252 – 474	304 – 556	260 – 418	238 – 460	370 – 528
		significance	$p < .0005$	$p < .0001$	$p < .0017$	$p < .0007$	$p < .0001$	$p < .001$
	Positive clusters	time(ms)	482 – 894	502 – 876	Not sig.	444 – 998	500 – 838	Not sig.
		significance	$p < .0001$	$p < .0001$	Not sig.	$p < .0001$	$p < .0002$	Not sig.
Logicians	Negative clusters	time (ms)	Not sig.	252 – 504	256 – 498	244 – 430	242 – 464	330 – 520
		significance	Not sig.	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0011$
	Positive clusters	time(ms)	Not sig.	550 – 898	542 – 998	454 – 908	522 – 892	810 – 998
		significance	Not sig.	$p < .0003$	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0035$

Discussion

The results of our experiment confirmed the main hypothesis that the violation of the scalar implicature can elicit an N400 effect similar to the N400 effect characteristic for standard semantic violations. In the case of comparing a clear semantic violation (*SF*) with the baseline condition *ST*, a typical N400 effect was observed for all subjects and for both sub-groups separately. This effect was additionally followed by a significant PNP. However, the N400 and PNP effects observed in the comparisons of the pragmatically infelicitous sentences (*SI*) with the baseline true (*ST*) or control false (*SF*) sentences, were dependent on on subjects' (behavioral) evaluation of the infelicitous sentences. These correlations of the size of the N400 effect and the PNP in the critical comparisons *SI* vs. *ST* and *SF* vs. *SI* and with the pragmatic vs. logical evaluation of underinformative sentences allows us to conclude that these effects were triggered by a pragmatic violation only if a person explicitly adopted the implicature as a part of the sentence's truth-conditions.

The interesting fact is that our participants were almost evenly divided into "pragmatists" and "logicians". This result could be interpreted as indicating an ad hoc semantic decision taken by a subject who had to choose between the two possible interpretations of the same linguistic expression. This conjecture is further reinforced by the fact that there were no significant differences between the two groups: neither involving their ERP differences in the control comparisons, their age, gender, AQ quotient, nor results from any part of the intelligence test. Thus, one can conclude that the recorded ERP effects in the critical comparisons *SI* vs. *ST* and *SF* vs. *SI* indicated primarily the person's truth-value evaluation of the underinformative sentences.

The effect of truth-value judgment vs. priming

A significant N400 effect was observed in all cases in which the compared conditions differed with respect to the truth-value judgment given by subjects. However, comparing two *false* conditions *AF2* vs. *AF*, and *SF* vs. *SI* in the case of "pragmatists", also resulted in a significant N400 effect, even though in both these cases the compared conditions did not differ with regard to the given truth-value evaluation. The N400 effect in the regarded comparisons can be associated with priming: In conditions $AF2$ and EF the critical words denoted a completely new object which was not depicted in the respective model, whereas in conditions *AF* and *SI* the critical word referred to one of the depicted objects. Thus, even though in both compared conditions the sentences were judged as false, only in *AF2* and *SF* the critical word was not visually primed by the preceding model. This lack of priming gave rise to more negative amplitudes in the N400 time-widows. Interestingly, the modulation of the PNP through priming seemed marginal and of a different sort. There was no significant cluster for "pragmatists" in the comparison *SF vs. SI*. The clusters found for the comparisons *AF2* vs. *AF* (not significant for "pragmatists") were 200*ms* delayed with respect to the PNPs in other comparisons and had a different topographical distribution, i.e. a more left-hemispheric extension. Thus, a clear centro-parietal PNP with an onset around 500*ms* and lasting for several hundred *ms* occurred only for comparisons in which there was a difference in the truth-value evaluation.

Conclusion

A strong version of the default theory cannot be supported by our results. If the implicature was by default incorporated into the sentence's meaning at the early stage of semantic processing, then we could expect an indication of the recognition of the implicature violation (e.g. the N400 effect) irrespectively of the subject's final decision to cancel this implicature. However, we found such an indication, i.e. the N400 effect, only when the implicature was adopted as part of the sentence's truth-conditions. An alternative approach according to which the implicature occurs as a purely context-based inference seems neither fully supported by our data. If the scalar implicature is based on post-propositional inferential processes, whereas it is the semantic meaning that is the default interpretation of *some*, why would the violation of this implicature trigger any N400 effect in the first place? In that case we would rather expect only a P600 effect reflecting some additional sentence reprocessing.

We find our results convincing enough to shed doubt on the well-established default- vs. context-based distinction as oversimplified and not providing sufficient theoretical framework to explain the nature of scalar implicatures. This leaves space for a more refined theory that would take into account both: a possibility of an incremental integration of the implicature into the sentence meaning as well as a non-mandatory and non-automatic character of the implicature. One theory of this sort could be the truth-conditional pragmatics account, which seems to be well-supported by our data. The correlation between the truth-value responses and the recorded ERP effects suggests that the implicature was incorporated into the sentence's truth-conditions at the early stage of semantic processing. Thus, the scalar implicatures can become a part of the truth-conditional content. Having said this we are aware that it is not the only way to account for the presented data. A more in-depth philosophical discussion is still needed for which the limits of this paper do not allow.

Acknowledgments

This study was supported by the *Stiftung Mercator* Foundation within the project *Theory of Memory*.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5-17.

Borg, E. (2012). *Pursuing meaning*. Oxford University Press.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar implicatures. *Journal of Memory and Language*, *51*(3), 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalized implicatures generated by default? An on-line investigation into the role of context in generating pragamtic inferences. *Cognition*, *100*(3), 434-463.

Cappelen, H., & Lepore, E. (2005). A tall tale: In defense of semantic minimalism and speech act pluralism. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning, and truth.* (p. 197219.). Oxford: Oxford University Press.

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *An international handbook of natural language meaning.* Mouton de Gruyter.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (p. 11-42). Georgetown University Press.

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463-470.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Levinson, S. C. (1983). *Pragmatics* (B. L. Catalounia, Ed.). Cambridge University Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.

Nieuwland, M., Diman, T., & Kuperberg, G. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324-346.

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

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2012). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research*, *1490*, 134-152.

Recanati, F. (2010). *Truth-conditional pragmatics*. Oxford University Press.

Van Berkum, J., Koornneef, A., Otten, M., & Nieuwland, M. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158-171.

Van Petten, C., & Luka, B. J. (2012). Review: Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190.