ERP indices of situated reference in visual contexts

Elli N. Tourtouri, Francesca Delogu, and Matthew W. Crocker
{elli, delogu, crocker}@coli.uni-saarland.de
Department of Computational Linguistics,
Saarland University, Saarbrücken, Germany

Abstract

Violations of the maxims of Quantity occur when utterances provide more (over-specified) or less (under-specified) information than strictly required for referent identification. While behavioural data suggest that under-specified expressions lead to comprehension difficulty and communicative failure, there is no consensus as to whether over-specified expressions are also detrimental to comprehension. In this study we shed light on this debate, providing neurophysiological evidence supporting the view that extra information facilitates comprehension. We further present novel evidence that referential failure due to under-specification is qualitatively different from explicit cases of referential failure, when no matching referential candidate is available in the context.

Keywords: informativity; over-specification; under-specification; referential processing; ERPs

Introduction

The gricean maxims of Quantity stipulate that speakers should provide no less (first maxim) and no more (second maxim) information than required for the purposes of the exchange. In visually situated communication, violations of these maxims occur through the use of referring expressions that provide less (under-specified) or more (over-specified) information than strictly required (minimally-specified) for the identification of the target referent. For example, the use of the expression "the yellow bowl" is over-specified in a context with only one bowl present (the colour modifier is superfluous), under-specified in a context with two yellow bowls (the modifier does not disambiguate between the two objects), while it is minimally-specified when a second bowl that differs in colour is co-present (the colour adjective is necessary and sufficient for identification).

Although it has not yet become clear why speakers may choose to include excessive information in their referring descriptions, empirical data demonstrate that they do so quite frequently, while they very rarely provide less information than necessary (Deutsch & Pechmann, 1982; Engelhardt, Bailey, & Ferreira 2006; Ferreira, Slevc, & Rogers, 2005; Nadig & Sedivy, 2002; Pechmann, 1989, inter alia), suggesting that speakers are not (fully) gricean. On the other hand, there is no clear evidence regarding the online sensitivity of listeners to the gricean maxims, that is, whether or not violations of the maxims of Quantity on behalf of the speaker result in processing difficulty for the listener. While some offline studies provide support to the intuition that under-specification impairs comprehension (Davies & Katsos, 2013, experiments 1 & 2; Engelhardt et al., 2006, experiment 2), there is no conclusive evidence as to how, if at all, over-specification affects processing. A number of studies suggest that, if not beneficial to comprehension, the inclusion of extra information is at least as good as minimal specification (Arts, Maes, Noordman, & Jansen, 2011; Arts, Maes, & Noordman, 2004), while others advocate that over-specification leads to impairments in comprehension (Engelhardt et al., 2006, experiment 3; Engelhardt, Demiral, & Ferreira, 2011; Davies & Katsos, 2013).

In an ERP experiment, Engelhardt, Demiral and Ferreira (2011) used the referential processing task to present participants with two-object visual displays concurrent with audio instructions to look at the target. In one display, objects were either of the same type, e.g. two stars – rendering modification of the noun necessary for target identification – or of different types, e.g. a star and a circle – when the mention of only the head noun would be sufficient. The experiment used a 2x2 design crossing display (same/different) and modifier (colour/size). An N400-like effect was found for the adjective in response to over-specified relative to minimally-specified expressions, while behavioural data suggested that it took longer for participants to identify the target object upon hearing an over-specified description (different-object display). The findings were interpreted as an indication that the inclusion of additional information in referring expressions is harmful for comprehension. However, we believe that this might be a too strong conclusion, especially since over-specified instructions always accompanied different-object displays, possibly raising a confound, namely that any effect might simply be due to the (slightly more complex) display type, and not to the inclusion of extra information in the instruction. What is more, these results might merely reflect that extra information is strikingly redundant when visual context is highly simplified (as the two-object scenes used here), while different processes may be at play in the presence of more demanding visual settings. Furthermore, one might expect any effects of over-specification – beneficial or detrimental – to occur (also) at the noun, and it is not clear why Engelhardt and colleagues only focused on the adjective region.

In this paper we present an ERP study that tested the effects of over- and under-specification on language comprehension in visually-situated settings. We employed a referential processing task similar to the one in Engelhardt et al. (2011), but, crucially, with visual scenes that were more complex and accommodated all conditions. Furthermore, we tested the intuition that under-specification is detrimental to language understanding – as it leads to

referential failure – by comparing it to explicit cases of referential failure when no matching referential candidate is available in the visual context.

Experiment

In an ERP experiment participants listened to instructions like "Find the yellow bowl" in German, paired with four visual contexts such as the ones in Figure 1. We contrasted "the yellow bowl" in B, where the noun alone is sufficient for target identification (OS), and C, where the adjective does not help disambiguate (US), to A, where the adjective is necessary and sufficient (minimally-specified, MS). A mismatch (MM) condition served as a case of explicit referential failure, where the adjective and noun were both represented in the display but by different objects (D).

Given that over-specification is ubiquitous in language use, we hypothesised that OS would be facilitatory, or at least as good as MS, as speakers would unlikely use redundant information if this hindered comprehension for their listeners. As for US, our hypothesis was that it would be detrimental to comprehension, but possibly yielding a qualitatively different effect than MM, since US leads to an unresolved ambiguity, while MM raises a question as to the validity of the information provided by the adjective and noun.

Method

Participants Thirty-three Saarland University students (mean age 25, 11 male) participated in the experiment and were monetarily compensated for their participation. They were all right-handed native speakers of German with normal or corrected-to-normal vision and no problems with colour perception.

Materials For creating the visual stimuli we used pictures of 30 everyday objects that differed along the dimensions of type (e.g., bowls, mugs, etc.), colour (red, blue, green, yellow) and pattern (dotted, striped, checkered). We opted for colour and pattern as distinguishing features, since they are both intrinsic to the objects (therefore, not requiring comparison, as in the case of relative features, such as size). Colour hue and brightness were adjusted using GIMP (Version 2.8.10). In order to make sure that not only objects were identifiable in all colours and patterns, but also that the descriptions to be used in the experiment would not diverge from participants' naming preferences, we conducted an offline picture naming study. We presented 24 independent participants with the object images in all colours and patterns (distributed over 8 lists), and asked them to name the objects including a colour and pattern term. Only objects with naming agreement above 80% were used to create the visual stimuli.

A set of 128 items was created. Each item comprised one spoken instruction (containing either a colour or a pattern description) and four displays (essentially four versions of the same display counterbalancing the target position within the item, and the colours and patterns per object type throughout the experiment). Crucially, experimental displays were constructed so as to accommodate all four types of descriptions, so that the display would not reveal the condition. To this end, six objects were necessary per display: Two same-object pairs for the MS and US conditions, and two unique objects for OS and MM. The objects were arranged in an oval-shaped array, as shown in Figure 1. Because the determiner in German is marked for gender, visual displays employed only same-gender objects, in order to assure that the target referent would not be revealed before the adjective. It was also taken care that none of the nouns represented in a display would begin with the same phoneme, so as to make sure that disambiguation would always occur at noun onset.

In total, 640 visual displays were created, of which 512 were experimental (128 x 4 versions), 128 were fillers and another 12 were for practice. Half of the displays were designed to combine with colour descriptions and the other half with pattern descriptions. In the colour displays (cf. Fig.1) the MS pair (the two bowls in A) shared pattern, but differed in colour. The OS referent (the bowl in B) was of unique colour, but not unique pattern. The US objects (the two bowls in C) shared colour, but differed in pattern, and the MM target (the bowl in D) was of unique pattern, but not colour. Pattern displays were created following the same set-up, only objects that shared colour in the colour displays would now share pattern, and so on.

This resulted in an apparent inconsistency between colour and pattern displays; i.e., there were 4 colours in the first, while only 3 in the latter. We counterbalanced that in the fillers, by coupling the 3-colour displays with colour instructions and the 4-colour displays with pattern instructions. The target position was also counterbalanced in the fillers, where the target could occupy any of the 6 positions. What is more, it was taken care that across all displays – experimental and filler – the target objects occupied each of the 6 positions an equal number of times.

The visual displays were paired with German instructions like "Finde die gelbe Schüssel" (Find the yellow bowl) for the displays in Figure 1. All instructions started with the same two words "Finde den/die/das…", continued with a colour or pattern pre-nominal adjective and finished with the head noun. Filler instructions differed in that they contained one, two or zero modifiers, thus rendering all filler descriptions minimally-specified. Audio stimuli were recorded with neutral intonation by a young, female speaker of German, in a sound-proof recording booth using Cubase AI 5. As speech was continuous (there was no attempt to insert pauses in between words), recordings were then annotated for adjective and noun onsets using Praat (Version 5.3). The mean duration of the adjective was 481.3ms (SD = 32ms), and that of the noun was 557.2ms (SD = 75.7ms).

Four lists were created using the Latin square design. Lists were pseudo-randomized so that no more than two experimental items were consecutive at any point in the list, and that, even when a filler intervened, there would not be

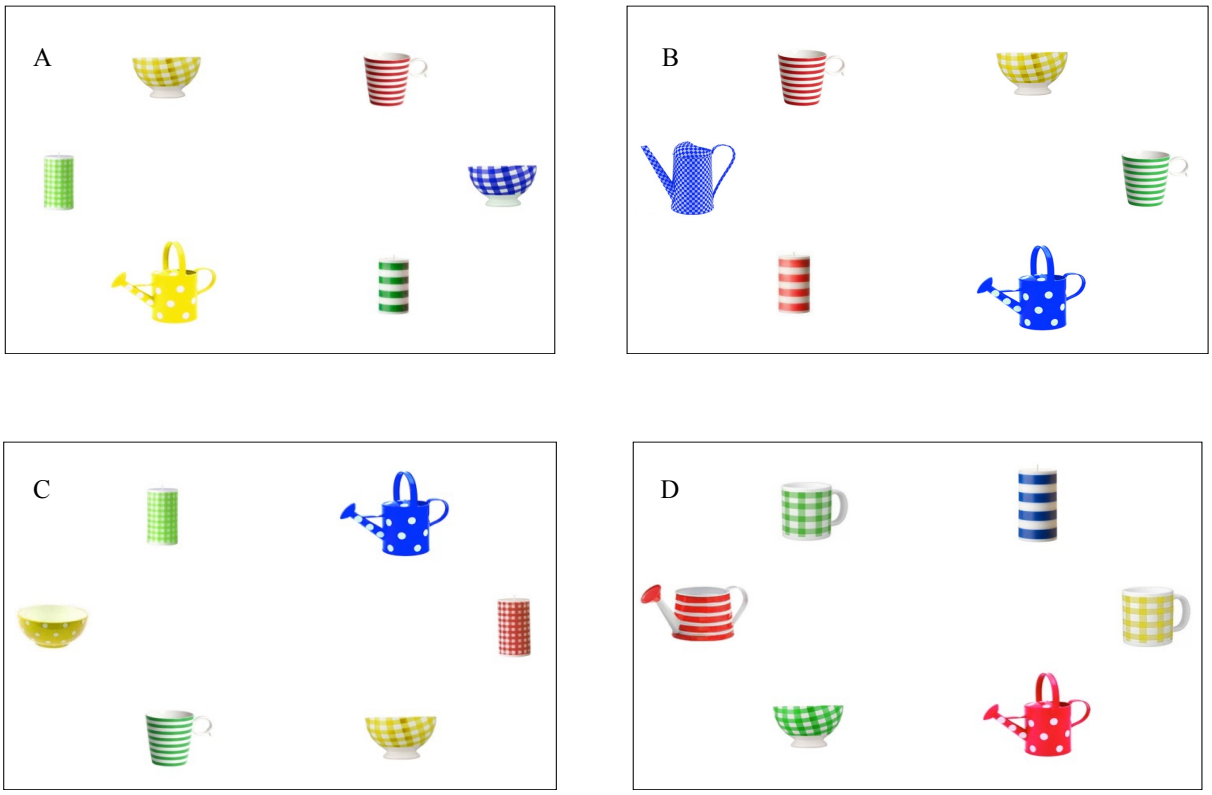

Figure 1. Sample displays for a colour item. All four displays are paired with the instruction "Find the yellow bowl", resulting in the four conditions: Minimally-specified (MS) in A, over-specified (OS) in B, under-specified (US) in C, and mismatch (MM) in D.

two items of the same condition in a row. Stimuli were presented in 8 blocks of 32 trials each, and block presentation order was also counterbalanced. An additional block of 12 filler trials was used for practice.

Based on the findings by Engelhardt et al. (2011), we predicted that over-specification would modulate an N400-like component, with OS being less negative than MS, since the adjective renders the noun predictable. For US compared to MS we expected a component related to processing difficulty that might differ qualitatively from that yielded in MM, due to the differing nature of referential failure in the two conditions.

Procedure The experiment was implemented and run using the E-prime software (Psychology Software Tools, Inc.). Participants were seated alone in a sound-isolated and electromagnetically shielded cabin. Displays were presented on a 1680 x 1050 resolution monitor.

Following a 3s preview, a cross appeared in the middle of the screen that participants had to fixate, and 500ms later the audio instructions were played. The display remained on the screen without the fixation cross for another 500ms after the instructions. Next, participants were prompted to carry out the task, which was to indicate which side of the display the target object appeared on (MS and OS conditions), or whether such a decision was not possible (US and MM conditions) by pressing the corresponding button on a button box in front of them as quickly and accurately as possible.

The EEG was recorded from 26 Ag/AgCl electrodes placed on the scalp according to the standard 10-20 system and the signal was amplified by a BrainAmps DC amplifier (Brain Products). Eye movements and blinks were monitored by electrodes placed on the outer canthus of each eye, and above and below the right eye. Impedances were kept below 5kΩ. The EEG signal was digitized at a sampling rate of 500Hz and re-referenced offline to the average of both mastoid electrodes.

Analysis The EEG signal was filtered offline (30Hz high cut-off). Single-participant averages were then computed in a 1000ms window per condition relative to the onset of the adjective ("yellow") and head noun ("bowl"), and aligned to a 200ms pre-stimulus baseline. Trials were semi-automatically screened offline for eye movements, blinks, electrode drifts, and amplifier blocking. After artefact rejection 8 participants with less than 18 trials left were excluded from analyses. Only artefact-free ERP averages time-locked to the onset of the critical regions entered the analyses. We performed omnibus repeated measures

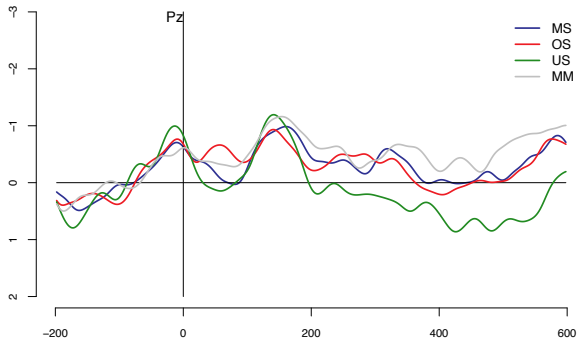

Figure 2. Averaged ERPs for the four conditions time-locked to the adjective onset (vertical line) at electrode Pz.

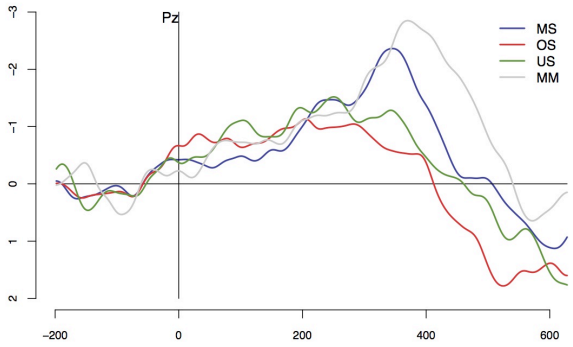

Figure 4. Averaged ERPs for the four conditions time-locked to the noun onset (vertical line) at electrode Pz.

ANOVAs on mean amplitudes crossing Informativity (4 levels) with an Electrode factor (17 levels). Any effects and interactions were followed up with separate pairwise comparisons (OS, US, MM vs. MS, and US vs. MM).

Results

Reaction time analyses revealed that participants were faster ($p < .01$) to identify the target in the OS (439ms, SD = 250.45), and slower ($p < .01$) in the MM (540ms, SD = 348.84) compared to the MS (480ms, SD = 306.37) condition. Interestingly, the comparison between the US (482ms, SD = 357ms) and MS conditions did not result in a significant difference ($p > .05$). Response accuracy was high overall; participants pressed the correct button over 90% of the time in the OS, MS and MM conditions and 86.80% of the time in the US condition.

Visual inspection of the ERP waveforms time-locked to the adjective (Fig.2) shows a larger positivity for US compared to MS starting after 200ms and reaching maximum after 400ms. The omnibus ANOVA between 400-600ms yielded a significant interaction of Informativity x Electrode ($F(48,1200) = 1.57$, $p = .008$). This effect was further explored with pairwise comparisons. The comparison between US and MS revealed a marginal effect of Informativity ($F(1,25) = 3.14$, $p = .088$). As shown in Figure 3A the effect was broadly distributed and slightly more pronounced on the right electrode sites. The comparison between US and MM yielded a significant Informativity x Electrode interaction ($F(16,400) = 2.88$, $p < .001$) and Figure 3B displays the distribution of this effect. None of the other comparisons reached significance. These results indicate that something fundamentally different is going on in US compared to any of the other conditions, crucially also compared to the explicit case of referential failure (MM). This is possibly the fact that it is only in the US condition that participants are able to identify the target category (e.g., the bowl) already by hearing the adjective ("yellow") as the instruction unfolds, but, crucially, also fail to pin down the target object. We will return to this point in the discussion.

The ERPs time-locked to the noun (Fig. 4) show a graded negativity peaking around 400ms, with MM being the most negative and OS the least negative. The omnibus ANOVA in the 300-500ms time-window revealed a main effect of Informativity ($F(3,75) = 6.23$, $p < .001$) and an Informativity x Electrode interaction ($F(48,1200) = 2.66$, $p < .001$). Pairwise comparisons for MM vs. MS yielded an interaction of Informativity x Electrode ($F(16,400) = 5.08$, $p < .001$). As shown in Figure 3C the distribution of the effect is centro-parietal, which is the typical distribution of the N400 effect. The comparison between OS and MS in the same time-window revealed a main effect of Informativity ($F(1,25) = 8.26$, $p = .008$), and again a centro-parietal distribution (Fig. 3D). The timing and topographic

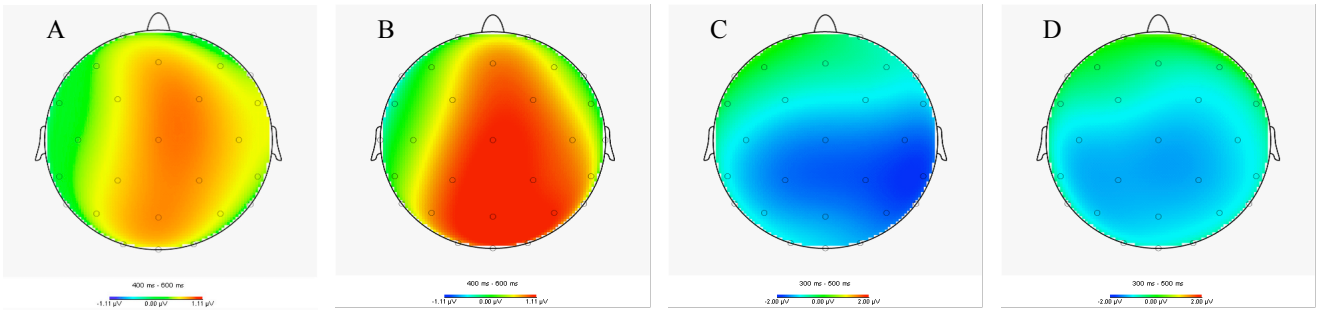

Figure 3. Topographic maps showing the effects of US minus MS (A) and US minus MM (B) in the 400-600ms time-window post-adjective onset. C is showing the effect of MM minus MS in the 300-500ms time-window post-noun onset, and D the effect of MS minus OS in the 300-500ms time-window post-noun onset.

distribution of the two effects indicates that Informativity at the noun modulates the N400 component. While the comparison between US and MS in the 300-500ms time-window did not reach significance,[1] the US vs. MM comparison revealed a main effect of Informativity ($F(1,25) = 10.53$, $p = .003$) and an Informativity x Electrode interaction ($F(16,400) = 3.19$, $p < .001$). While this finding seems to corroborate the view that referential failure due to under-specification is different from referential failure due to mismatch, we refrain from drawing strong conclusions from the effects (or lack thereof) elicited by the US condition in the noun region (see Footnote 1).

General Discussion

In this study we investigated the influence of over- and under-specified expressions on visually-situated referential processing. The results that we report offer two important insights: Firstly, we provide evidence that over-specification is beneficial rather than detrimental to language comprehension, as indexed by the decreased negativity elicited for the OS relative to the MS condition. Secondly, we tested the intuition that under-specification impairs comprehension as it leads to referential failure, by contrasting it, not only to minimal-specification, but, importantly, to cases of explicit referential failure (mismatch), and showed that the two processes result in qualitatively different effects.

Specifically, while at the adjective region there were no differences between the MM, OS and MS conditions, we found a graded centro-parietal negativity peaking around 400ms after the onset of the noun for MM, MS, and OS, where MM was the most and OS the least negative. We interpret this N400-like effect as reflecting the predictability of the noun,[2] as determined by the visual context in combination with the information provided by the adjective. That is, while in the MS condition the adjective provides enough information to help narrow down the referential space to two objects (cf. the yellow bowl and the watering-can in Fig.1A), with the noun disambiguating, then, between the two, in both the OS and MM conditions the adjective ("yellow") can already single out the target referent (the bowl and mug, in Fig.1B and 1D, respectively), thereby raising specific expectations for the noun. When these expectations are disconfirmed (MM condition), the N400 elicited by the noun is higher than it is in the MS condition, while when they are confirmed (OS condition), the N400 amplitude is lowest. This last finding is at odds with the results of Engelhardt et al. (2011), as it suggests that not only is over-specification not detrimental to comprehension, but that it is in fact beneficial, at least when in presence of a complex visual context. Such an interpretation is corroborated by the reaction time data, which were faster in the OS, while slower in the MM, in comparison with the MS condition.

As for the cases when the information given is less than minimally required for target identification, the pattern of results first and foremost reveals that referential failure due to US is qualitatively different than that in MM. While the ERPs for the OS, MS and MM conditions overlap throughout the adjective region and start diverging only at 300ms after the noun onset, there was a significant positive deflection for US compared to MS already at the adjective region (cf. Fig. 2). What is so unique about the US condition that is reflected in these findings? Crucially, while in both the US and MS conditions the adjective picks out two out of the six objects, in the MS condition these are of a different type (cf. yellow bowl and watering-can in Fig.1A), and the noun is still required for disambiguation. In other words, in the MS condition, both the adjective and the noun are necessary to fully disambiguate the target, and each of them provides information that incrementally restricts referential space. By contrast, in the US condition the adjective identifies exactly two objects that are of the same type (cf. the two yellow bowls in Fig.1C), rendering the adjective redundant. What is more, processing of the adjective in the US condition is different than in the OS and MM conditions, as well: Even though the adjective in OS and MM is also redundant, it does however help single out the target referent already before any information about its type comes in, giving rise to predictions about the head noun,[3] and potentially making processing easier and faster (OS condition). In the US condition, on the other hand, the adjective is not only redundant, but it is also unhelpful, since the two objects it pins down are of the same type, giving away the head noun, but at the same time having listeners await for more information, as they discover that the upcoming noun is not going to help disambiguate. We believe that it is exactly this realisation that is reflected in this positivity.

The influence of providing less information than communicatively necessary on listeners' brain responses during comprehension is, to our knowledge, under-investigated. One study that touches on this issue, however, provides results consistent with our current findings. Hoeks, Stowe, Hendriks, and Brouwer (2013) investigated the processing of partial answers to questions. They had participants read short dialogues that comprised questions like "What did the mayor and the alderman do", and responses that only answered half the question and left information about the other half pending ("The mayor praised the councilor"), as their brain responses were measured. Relative to a neutral condition where the question was general ("What happened") – and the answer was,

[1] Note, however, that the pre-stimulus baseline correction was performed on an interval displaying a significant difference between US and MS (the last 200ms of the adjective). This may have artificially pulled the two waveforms together, thereby masking any potential effect of US vs. MS in the noun region.

[2] Cf. Kutas and Federmeier (2011) for a review on the N400 as indicator of predictability (even though the literature so far does not extend to situated language processing).

[3] Although these predictions turn out to be inaccurate in the MM condition, they are incrementally received as helpful (cf. the discussion about the graded negativity, above).

therefore, complete – partial answers resulted in a broadly-distributed positivity, which started around 350ms after the onset of the critical word ("councilor") and lasted through the 600-900ms time-window. The authors interpret this positivity as reflecting increased effort in updating or reorganising a representation of what is being communicated. Analogously, the positivity elicited by US compared to MS at the adjective possibly reflects, on the one hand, the realisation that this information is not helpful – since it picks out two objects of the same type – and, on the other, some process of updating the mental model of what is being said. This update can amount to the general expectancy for disambiguating information to come in before the noun (which is already predicted), or even the formulation of specific predictions as to what the next adjective should be, given information so far (cf. "dotted" or "checkered" in Fig.1C).

Future work is necessary to interpret the effect of under-specification and determine whether this positivity indeed reflects processes of updating the listener's mental representation. For example, under this account, if a third object of the same type but different colour was introduced in the visual context (e.g., a red bowl in Fig.1C), one should not expect a positive deflection after "yellow", since the adjective now rules out one object and helps narrow down the referential space.

Conclusions

Our findings demonstrate that ERPs index the full spectrum of situated referential processes, offering two important insights. Firstly, we observed N400 sensitivity to the (visually-determined) predictability of the noun in the MM, MS and OS conditions, suggesting that over-specification is not detrimental, but rather beneficial to language comprehension. Secondly, we show that listeners rapidly identify unhelpful information: The adjective in the US condition fails to distinguish between the two objects of the same type, resulting in a positive deflection relative to both the MS and MM conditions. This effect indicates that referential failure due to under-specification is qualitatively different from explicit cases of referential failure (mismatch).

Acknowledgments

We thank Zijian (Fabio) Lu, Torsten Jachmann and Yoav Binoun for helping with experiment implementation and data collection, and Vera Demberg and Harm Brouwer for useful discussions. This research was supported by the "Multimodal Computing and Interaction" Cluster of Excellence and by SFB1102 "Information Density and Linguistic Encoding" awarded by the German research foundation (DFG).

References

Arts A., A. Maes, L. Noordman, & C. Jansen (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43*, 361-374.

Davies, C., & N. Katsos (2013). Are speakers and listeners 'only moderately Gricean'? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics, 49(1)*, 78-106.

Deutsch W., & T. Pechmann (1982). Social interaction and the development of definite descriptions. *Cognition, 11*, 159-184.

Engelhardt, P. E., K. Bailey, & F. Ferreira (2006). Do speakers and listeners observe the Gricean Maxim of Quantity. *Journal of Memory and Language, 54, 554-573*.

Engelhardt P. E., Ş. B. Demiral, & F. Ferreira (2011). Over-specified referring expression impair comprehension: An ERP study. *Brain and Cognition, 77*, 304-314.

Ferreira, V. S., L. R. Slevc, & E. S. Rogers (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263-284.

Grice, P. (1975). Logic and conversation. In: P. Cole & J. Morgan (Eds.) *Syntax and Semantics: Speech Acts* (Vol.III), pp. 510-516. New York: Academic Press.

Hoeks, J. C. J., L. A. Stowe, P. Hendriks, & H. Brouwer (2013). Questions Left Unanswered: How the Brain Responds to Missing Information. *PLoS ONE, 8(10)*: e73594. doi:10.1371/journal.pone.0073594.

Kutas M., & K. D. Federmeier (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP), *Annual Review of Psychology, 62*, 621-647.

Maes A., A. Arts, & L. Noordman (2004). Reference management in instructive discourse. *Discourse Processes, 37*, 117-144.

Nadig A. S., & J. C. Sedivy (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychological Science, 13(4)*, 329-336.

Pechmann, T. (1989). Incremental Speech Production and Referential Overspecification. *Linguistics, 27(1)*, 89-110.