Specificity and entropy reduction in situated referential processing

Elli N. Tourtouri (elli@coli.uni-saarland.de)
Francesca Delogu (delogu@coli.uni-saarland.de)
Matthew W. Crocker (crocker@coli.uni-saarland.de)

Department of Language Science & Technology
Saarland University, Saarbrücken, Germany

Abstract

In situated communication, reference to an entity in the shared visual context can be established using either an expression that conveys precise (minimally specified) or redundant (over-specified) information. There is, however, a long-lasting debate in psycholinguistics concerning whether the latter hinders referential processing. We present evidence from an eye tracking experiment recording fixations as well as the Index of Cognitive Activity – a novel measure of cognitive workload – supporting the view that over-specifications facilitate processing. We further present original evidence that, above and beyond the effect of specificity, referring expressions that uniformly reduce referential entropy also benefit processing.

Keywords: referential processing; over-specification; visual entropy reduction; eye tracking; Index of Cognitive Activity

Introduction

Grice's maxims of Quantity (Grice, 1975) stipulate that speakers' utterances be minimally informative, avoiding redundancy. In visually situated communication, this predicts utterances should provide strictly the information necessary for the identification of a referenced object. For example, in the context of a blue and a green ball, the adjective "blue" is necessary to unambiguously establish reference. When there is only one ball, however, the adjective becomes superfluous. Such over-specifications – expressions that convey more information than minimally required – are, however, produced by adult speakers at an estimated rate of 10-60% (see Engelhardt, Bailey & Ferreira, 2006, and references therein).

Even though Grice arguably did not intend to make any implications about the cognitive processes associated with the violation of his maxims (cf. Geurts & Rubio-Fernández, 2015), over the past few decades, psycholinguistic research has tried to test their empirical validity. It remains under debate, however, whether or not over-specifications are detrimental to referential processing. A number of studies have suggested that over-specifications impair listeners' online processing and lead to slower and less accurate identification of the target (e.g., Engelhardt, Bailey & Ferreira, 2006; Engelhardt, Demiral & Ferreira, 2011; Davies & Katsos, 2013), while others find evidence that they are as good as minimal descriptions or may even facilitate processing (e.g., Arts, Maes, Noordman & Jansen, 2011; Tourtouri, Delogu & Crocker, 2015).

In an ERP experiment, Tourtouri, Delogu and Crocker (2015) presented participants with visual scenes of 6 objects and audio instructions to locate a target, like "*Find the yellow bowl*" (in German). The experiment manipulated the specificity of the referring expression by combining the same instruction with different visual displays that rendered it minimally or over-specified. An attenuated N400 effect was found on the noun for over- compared to minimally-specified references. This finding was interpreted as evidence that over-specifications are in fact beneficial to referential processing, at least when in the presence of visual displays where the over-specified adjective identified exactly one object. That is, at "*yellow*" the bowl was the only object that fit the description. Interestingly, both color and pattern adjectives were used to identify the target, and the effect was present for both types of adjectives, suggesting that any facilitation of over-specification is not merely due to the perceptual salience of color. It can be argued, however, that the reduced N400 for over-specifications may just reflect the predictability of the noun as determined by the information on the visual scene in combination with the linguistic input up to the adjective. Therefore, it still remains unanswered whether over-specification has a *general* facilitatory effect, even when displays allow the adjective to select a second object, which fits a minimally specified continuation of the referring expression, i.e., it is part of a contrast pair.

A similar question was addressed by Sedivy, Tanenhaus, Chambers and Carlson (1999) in a series of experiments that tested (among other things) whether intersective adjectives such as color are interpreted contrastively. Participants' eye movements were tracked while they heard instructions to manipulate objects in a workspace in front of them. The visual scenes consisted of four objects, two that formed a contrast pair differing in color, e.g. a yellow and a pink comb, and two singletons: one sharing the color feature with an object from the pair, e.g. a yellow bowl, and a distractor object of different color. The critical instruction mentioned either of the two objects with the shared feature, and was always heard second, following the instruction referring to an object from the contrast pair. An effect of referent type was found, such that if the target was part of the contrast pair it was looked at faster than if it was not. This result was taken to indicate that initially, before the noun was heard, listeners assigned a contrastive meaning to intersective adjectives, consistent with Grice's maxim of Quantity. We believe, however, that this may not be the case, especially since listeners' attention was already focused on the contrast pair, as the immediately preceding instruction always made reference to one of the contrasting objects (e.g., the pink

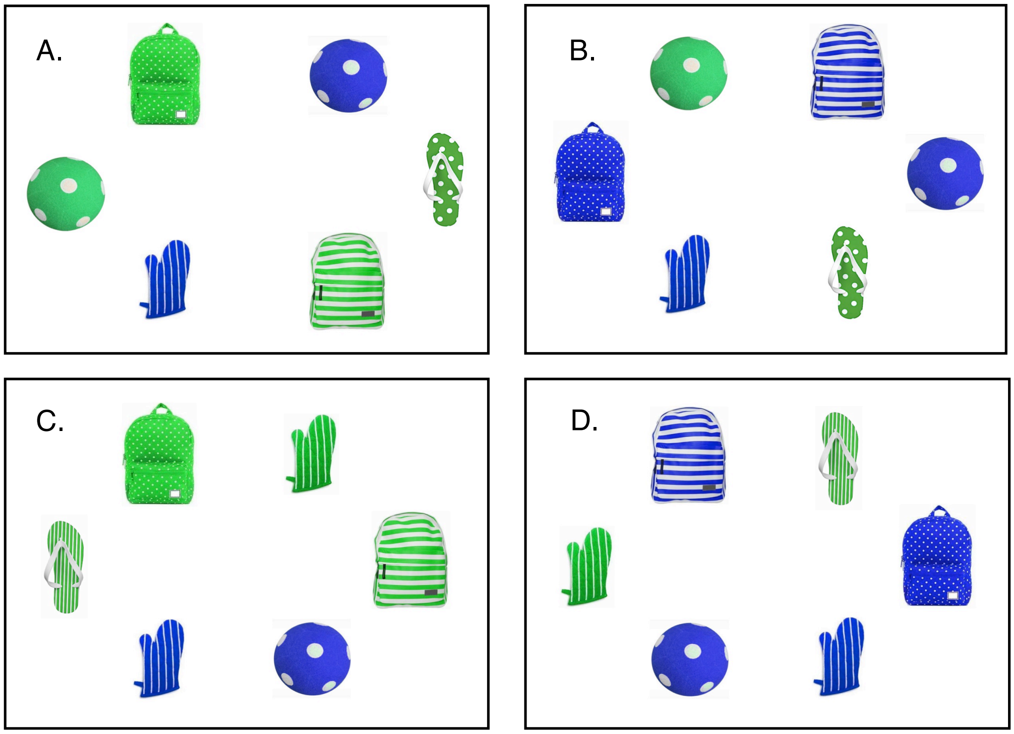

Figure 1. Sample visual stimuli for a color item, combined with the instruction "Find the blue ball". The four resulting conditions were A. Minimally specified – High reduction (MS-HR), B. Minimally specified – Low reduction (MS-LR), C. Over-specified – High reduction (OS-HR), D. Over-specified – Low reduction (OS-LR)

comb). Similar results were obtained in the subsequent experiments, where the critical instruction was heard first, but with the use of scalar adjectives, which inherently invoke comparisons between entities.

The current study seeks to determine whether and how over-specifications affect processing of pre-nominally modified referring expressions, when the visual context enables both a minimally and an over-specified reading of intersective adjectives, such as color and pattern. That is, how is referential processing influenced when the adjective is redundant (as in the bottom displays of Fig.1) as opposed to when it is required to uniquely identify the target (as in the top displays of Fig.1)? Furthermore, as the instruction sentence unfolds over time, incoming words incrementally restrict the set of referential candidates. Therefore, in situated communication, the information conveyed by a linguistic unit is determined by the extent to which it reduces the number of potential referents, in addition to the linguistic information of each word, as determined by its probability and preceding context (Shannon, 1948; Crocker, Demberg & Teich, 2016). In other words, the information on the word "*blue*" in the sentence "*Find the blue ball*" is not defined only in terms of its probability to occur in this (linguistic and visual) context, but also by the amount of uncertainty about the target (referential entropy) it reduces. For example, in the left-hand displays of Figure 1 "*blue*" reduces referential entropy by 1.58 bits, while in the right-hand displays it only reduces it by 0.58 bits. The noun, then, eliminates the remaining entropy, reducing it by 1 bit in the former and by 2 bits in the latter case, resulting in a uniform and a less uniform entropy reduction profile, respectively. This study also touches on whether, above and beyond any effects of specificity, the rate at which the linguistic input reduces referential entropy also influences processing. To examine these questions, we recorded participants' fixations as they viewed displays such as the ones in Figure 1, while listening to instructions like "*Find the blue ball*" in German, and present results from inspection probabilities to the objects of interest and the Index of Cognitive Activity (Marshall, 2000) per region.

The Index of Cognitive Activity

It is well established that fluctuations of the pupil size index cognitive effort in a variety of tasks, including language processing (e.g., Just & Carpenter, 1993). However, changes in the lighting conditions of the environment are also responsible for pupil dilation. The Index of Cognitive Activity (Marshall, 2000) is a measure of cognitive workload that separates variation in pupil size due to cognitive effort and due to light reflex, while also accounting for random noise. The small and rapid pupil dilations that remain are associated with higher cognitive workload (Marshall, 2002). Demberg and Sayeed (2016) showed, for example, that the ICA is sensitive to linguistic manipulations such as ungrammaticality, with conditions related to higher processing demands resulting in higher ICA values. They also demonstrated that ICA is particularly suitable for the Visual World Paradigm, since it is robust to the change of fixation positions and can thus complement the standard visual attention metrics in order to assess cognitive effort during linguistic processing.

Experiment

We used a 2x2 design crossing Specificity (Minimally specified vs Over-specified) and Entropy Reduction (Uniform vs Non-uniform). Based on findings that over-specifications are commonly used by adult speakers during production (cf. Engelhardt et al., 2006; Pechmann, 1989; Rubio-Fernández, 2016; Tarenskeen, Broersma & Geurts 2015), we hypothesized that over-specification would not impede referential processing, as rational speakers would unlikely use them so frequently if they did. We, therefore, expected that over-specified expressions (OS) would be as easy as, or easier than their minimally-specified (MS) counterparts (as found in the ERP study by Tourtouri et al., 2015). As for the entropy reduction manipulation, we generally expected a greater processing advantage in the uniform reduction (UR) compared to the non-uniform reduction (NR) conditions, as has been proposed for the related measure of surprisal (UID, Jaeger, 2010). Finally, we expected that the two factors should interact, namely that processing would be particularly benefited when the expression was OS and the redundant adjective contributed to the uniform reduction of entropy.

Method

Participants Twenty-four students from Saarland University (mean age 25, 7 male) participated in the experiment for monetary compensation. They were all native speakers of German with normal or corrected-to-normal vision and normal color perception.

Materials Pictures of 30 common use objects (e.g., balls, mugs, etc.) differing in color (blue, green and red) and pattern (checkered, dotted and striped) were employed to create the visual stimuli. Both color and pattern were used as distinguishing features, because they are intrinsic to the object, as opposed to scalar adjectives such as size that trigger comparisons to other entities on the display. This ensured that any looks to objects in contrast pairs would be driven due to the manipulation and not because of the adjective type. Furthermore, pattern was the mentioned property in half of the trials, in order to make sure that any effect of over-specification would not be merely due to color salience, but would be attributable to the experimental manipulation. Color hue and brightness were adjusted using GIMP (Version 2.8.10). Naming agreement was tested for the object pictures in an offline picture naming study to ensure that they were identifiable in all colors and patterns, and that the names used in the experiment matched participants' own naming preferences. Twenty-four independent participants were presented with the object images in all colors and patterns (distributed over 8 lists), and were asked to name them while always mentioning their colors and patterns. Only objects with a naming agreement of 80% or higher were employed in the visual stimuli.

A set of 120 experimental items was created, each item comprising one spoken instruction (with either color or pattern as the target feature) and four visual scenes (essentially four versions of the same scene). The target color, pattern and position were counterbalanced throughout the experiment. Displays for experimental items accommodated all four conditions for both target features, so that nothing would reveal the target before the instruction was heard. To this end, one visual scene contained 6 objects (two pairs of same-type objects and two singletons) in two colors and two patterns, such that the pairs made up the two MS and the singletons the two OS referents for both target features, as shown in Figure 1. Furthermore, displays never contained phonetic competitors (e.g., [ʃʏsəl] vs [ʃʏrtsə]), ensuring that disambiguation of the target would always occur on noun onset. For the same reason only same-gender objects were used per display, as German marks determiners for gender.

In total, 660 visual displays were created, of which 480 were used in experimental items (120 x 4 conditions), and 180 in fillers. Twelve of the fillers served as practice trials in a familiarization phase. Fillers differed from experimental items in multiple aspects. First, they differed in terms of their display structure, with almost half of the fillers depicting 4 objects (3 of the same type and one singleton) and the rest containing 6 objects. Six-object fillers either showed 2 contrast pairs and 2 singletons, where reference was always made to the contrast pair that was not relevant in the experimental items (e.g., the two rucksacks in Fig.1), or they showed a set of 3 same-type objects, a contrast pair and a singleton. The 3-object sets made a second modifier also required for target identification, thus adding more variation not only to the display types that participants viewed, but also to the referential entropy reduction possibilities. Secondly, fillers differed in terms of the specificity of their instructions, that could be minimally, over-, or under-specified (US), while care was taken so that throughout the entire experiment, participants would hear MS expressions to a greater extent than OS – as is the case in everyday language use – as well as a small portion of US. A set of fillers without pre-nominal modification was also used, that were essentially the minimally specified versions of the OS items, thus assuring that participants would not always expect to hear an adjective and that they would not get overly used to reference being redundant.

In experimental items, displays were paired with audio instructions containing a pre-nominally modified referring expression like "*Finde den blauen Ball*" (Find the blue ball) in Figure 1, that identified the target by mentioning either its color or its pattern. In fillers, instructions had zero, one or two modifiers. For the latter the order of mention of color and pattern adjectives was counterbalanced. Audio stimuli were recorded with neutral intonation by a young, female speaker of German, in a soundproof booth using Cubase AI5. Speech was continuous and no artificial pauses were inserted in between words. Sentences were then cut and annotated for adjective and noun onsets using Praat (Version 5.3). Mean word duration was 481.3ms (SD=32) for the adjectives and 557.2ms (SD=75.7) for the nouns.

Stimuli were distributed over 4 lists using the Latin Square design, and were pseudo-randomized for each participant. At least one filler appeared between consecutive experimental items, and items of the same condition did not appear more than two times in a row. Each participant saw 288 stimuli split in 4 blocks, which allowed for breaks in between blocks. Before the experiment started, a short practice session of 12 filler trials familiarized participants with the task. The experiment was implemented and run using E-prime 2.0 (Psychology Software Tools, Inc.).

Procedure An SMI RED500 eye tracker (SensoMotoric Instruments) attached to the bottom of a 25inch Dell monitor was used to track participants' eye movements at a rate of 250Hz. After they gave informed consent and read the instructions, participants were seated at a distance of approximately 60cm in front of the monitor using a chinrest to minimize head movements. They then completed a familiarization phase, during which the experimenter gave them feedback after each trial, ensuring that the task was clear before the experiment begun. Calibration was performed at the beginning of each block.

A trial started with a fixation cross appearing in the middle of the display for a period controlled by the experimenter. The objects then appeared while the cross was still on screen for another 500ms, and 1500ms later the audio instruction started. The objects stayed on the screen for another 500ms after the audio offset, and a prompt screen to the task appeared asking participants to indicate which side of the screen the target entity was on, or whether it was not possible to tell (US fillers) by pressing the corresponding button on a response pad in front of them. Displays were presented at a 1680×1050 resolution. One experimental session lasted on average 40min, depending on whether calibration had to be repeated.

Analysis For the analyses of both measures we considered the regions of the Adjective ("*blauen*"), and Noun ("*Ball*"). For the analysis of fixations, we compared inspection probabilities to areas of interest (AOI) across conditions. First, fixations shorter than 80ms were pooled with the immediately preceding or following fixation, if the distance between them was smaller than 12 pixels, otherwise they were excluded from the analysis. Subsequently, fixations to an AOI within a region, before a saccade outside the area was made, were counted as one inspection. For each AOI and region, we coded trials that contained at least one inspection to the AOI as 1, and trials that did not as 0. Therefore, mean values represent inspection probabilities per AOI and region.

As information about the target became incrementally available, different objects and different comparisons were interesting per region. In particular, at the Adjective, the only available information about the target was its distinguishing property, so the specificity manipulation is still irrelevant (it is still unknown whether the target is minimally or over-specified). We, therefore, compared inspections to the singleton and contrasting objects that bore the target property (cf. the blue ball & mitten in A&B, Fig.1) between UR and NR (collapsing across Specificity). Finally, at the Noun, when the target is revealed, both factors become relevant, so we contrasted inspection probabilities to the target (the blue ball: MS in A&B, OS in C&D, Fig.1) and to the competitor (the blue mitten: OS in A&B, MS in C&D, Fig.1) across conditions.

To calculate the ICA we used BeGaze™ with the ICA Module (SensoMotoric Instruments) and Workload^{RT} (EyeTracking, Inc.). Since the ICA values that the BeGaze™ software outputs are too coarse-grained for the type of effects we expect, we used the ICA Coefficients to compute ICA values per 100ms (see Demberg & Sayeed, 2016 for more details). Data points with a pupil diameter smaller than 2.5 SD of that participant were eliminated, and a mean ICA value for both eyes was calculated. As fixation positions are not relevant for the ICA, we were interested only in differences between UR and NR (collapsing across Specificity) for the Adjective, and across conditions for the Noun. We compared mean ICA values across conditions within a window of 600ms starting from the middle of each region. We analyzed inspection probabilities and ICA values using generalized linear mixed effects models (lme4 package, R Version 3.3.2) with random intercepts for participants and items, as well as random slopes for the predictors of interest. For the analysis of the Adjective, Reduction (UR vs NR) was the predictor of interest. For the analysis of the Noun, the models included the effects of Specificity (OS vs MS), Reduction (UR vs NR), and Target Feature (Color vs Pattern), and their interaction. When the maximal models did not converge, we simplified the random effects structure as suggested by Barr, Levy, Scheepers, and Tily (2013).

Results

Adjective

Singletons bearing the target feature (cf. the mitten in A&B, and ball in C&D) were inspected equally frequently in UR and NR (Coeff. = .083, SE = 2.317, Z = .829, p > .05). Contrast objects (cf. the blue ball in A&B), on the other hand, were more frequently inspected in UR than in NR (Coeff. = .329, SE = .0628, Z = 3.107, p = .001).[1] The ICA values (see Fig.2) did not differ significantly between UR and NR (Coeff. = -.031, SE = .0249, Z = -1.236, p > .05).

[1] Since in NR more entities bear the mentioned feature, and therefore attention is distributed across more objects, we do not take this difference to reflect any preference for a gricean/contrastive reading of the adjective.

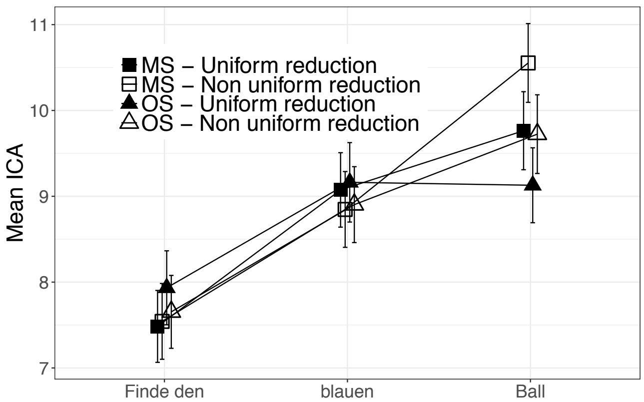

Figure 2. Mean ICA values per condition and region. Error bars represent 95% CI.

Noun

Analyses of inspection probabilities to the target (the blue ball) and competitor (the blue mitten) objects, including target feature as a predictor, produced a significant effect of feature (Coeff. = .205, *SE* = .104, *Z* = 1.971, *p = .048*). We followed up this effect with separate analyses for color and pattern items. For target inspections in color items (Fig.3), we found a main effect of Entropy Reduction with more inspections in UR than in NR (Coeff. = -.241, *SE* = .122, *Z* = -1.971, *p* = .048), as well as a marginally significant effect of Specificity (Coeff. = .237, *SE* = .127, *Z* = 1.877, *p* = .06), such that the target was inspected more frequently in OS than in MS. The analysis of inspections to the competitor resulted in a main effect of Entropy Reduction with the competitor receiving more inspections in UR than in NR (Coeff. = .39, *SE* = .151, *Z* = 2.583, *p* = .009), and no effect of Specificity (*p* > .05). For pattern items, none of the comparisons produced significant results (all *p* > .228). Interestingly, the ICA analysis produced main effects of both Entropy Reduction and Specificity for both color and pattern items (see Fig.2), such that ICA was higher for NR vs UR (Coeff. = -.07, *SE* = .024, *Z* = -2.96, *p* = .003), and for MS vs OS (Coeff. = .087, *SE* = .025, *Z* = 3.42, *p* < .001).

General Discussion

We investigated the effects of Specificity on situated language processing comparing listeners' inspection patterns and cognitive effort when exposed to minimally or over-specified reference. In accordance with previous research (cf. Arts et al., 2011; Tourtouri et al., 2015) we found a facilitation for OS vs MS on the noun, with the target object receiving more inspections when the referring expression included a redundant vs a contrastive adjective. However, this effect was observed only for color items, raising the question whether what facilitates processing is in fact color salience as opposed to over-specificity in general. The answer is provided by ICA, a novel measure of cognitive workload based on the count of rapid pupil dilations, which we used to directly assess the cognitive effort expended in each condition, showing that in both color and pattern items OS was indeed easier to process than MS. This discrepancy between the two measures seems to suggest that, while pattern is more difficult to perceive than color, its mention is nevertheless as beneficial to visual search as that of color. Further research is necessary to determine the relation between visual attention as measured by inspection probabilities, and cogntive load as measured by the ICA.

We further examined if and how processing is influenced by a more or less uniform reduction of referential entropy, i.e., of the size of the referential search space. Specifically, we contrasted conditions where the pre-nominal adjective reduced entropy from by 1.58 bits (UR) with cases in which entropy was decreased by only 0.58 bits (NR), to establish whether what determines efficient entropy reduction is determined by the more or less uniform decrease of entropy over the referential expression. Our results seem to provide evidence that processing is facilitated by the uniform reduction of referential entropy, though not in the predicted region. That is, in the Adjective there were no differences for either inspections (to singleton objects) or for ICA values between UR and NR. On the Noun, however, we found indications that the greater reduction of entropy at the first step contributing to a more uniform entropy reduction profile was preferred, since in (both MS and OS) UR conditions the target object collected more inspections (though only in color items), and, perhaps more interestingly, ICA values were lower than in NR (for all types of items). Importantly, this finding demonstrates that the ICA is sensitive to visual search difficulty, capturing differences in the cognitive effort expended with different rates of visual entropy decrease. With respect to our research question, visual search, and therefore referential processing, appears to be more efficient when the remaining set of possible referents at the final step is rather small, as is the case in UR. Interestingly, however, there is no penalty for this increased entropy reduction on the adjective.

We acknowledge that these results are open to alternative interpretations. For example, the absence of an Entropy Reduction effect on the adjective may be due to our operationalization of Uniform and Non-uniform Reduction, and not because the entropy reduction rate only affects the

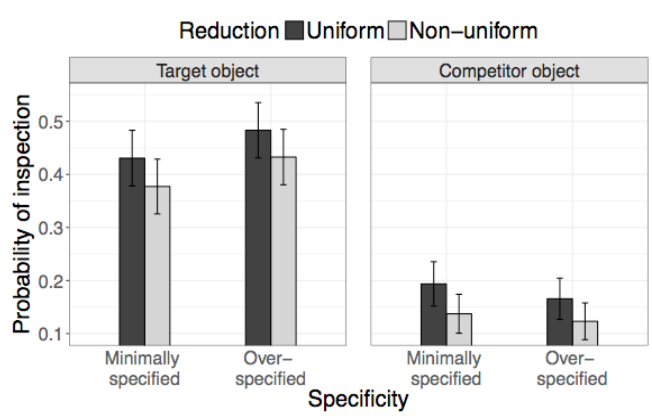

Figure 3. Inspection probabilities for the noun region in color items. Error bars represent 95% CI.

final step. That is, reducing referential entropy by 1.58 bits vs 0.58 bits may not be sufficient to induce a differential cost on the adjective. So, if the difference between the remaining entropy in the two conditions was enhanced, by using a larger referent set and going from e.g. 12 to 2 *versus* to 8 potential referents in Uniform and Non-uniform Reduction, respectively, might serve to amplify a reduction-related cost on the adjective in the Uniform condition. Relevant to this issue, it would be interesting to compare processing of OS as implemented in this experiment, with their MS counterparts, i.e. without modification of the noun (e.g., "*Finde den Ball*"), as the latter is not only a case of rapid reduction of entropy, but is also MS. Any facilitation for OS under this comparison should suggest that processing ease for OS is due to the insertion of an intermediate step in reducing visual entropy that makes reduction more uniform. A final possible explanation that is worth pursuing, as it is related to the nature of the ICA measurements, is that ICA may not be sensitive to such modulations of entropy reduction. In other words, ICA may only be able to capture whether visual search has been demanding or not. Future research is of course required to tackle these questions.

In sum, we present eye-tracking evidence confirming that the use of redundant noun modifiers (over-specification) facilitates referential processing. In addition, we showed that listeners rapidly exploit incoming information about the target to reduce the referential search space in situated comprehension. Greater reduction in referential entropy on the adjective – while not associated with any increase in cognitive load in that region – results in an overall more uniform entropy reduction profile and in reduced cognitive effort when processing the noun. This result leads us to conclude that efficient processing is determined by both the degree of specificity of the reference, as well as to the distribution of entropy reduction across the utterance.

Acknowledgments

We thank Mindaugas Mozuraitis and Maria Staudte for assistance with the eye tracking analyses, and Mirjana Sekicki, Vera Demberg and Ben Peters for help with the ICA analyses. This research was supported by SFB/CRC1102 "Information Density and Linguistic Encoding" awarded by the German Research Foundation (DFG).

References

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Crocker, Demberg, & Teich (2016). Information Density and Linguistic Encoding (IDeaL). *Künstliche Intelligenz*, *30*, 77-81.

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

Demberg, V., & Sayeed A. (2016). The Frequency of Rapid Pupil Dilations as a Measure of Linguistic Processing Difficulty. *PLoS ONE, 11(1)*. doi:10.1371/journal.pone.0146194

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

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

Geurts, B., & Rubio-Fernández P. (2015). Pragmatics and Processing*, Ratio, 28(4)*, 446-469.

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

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian journal of experimental psychology. 47(2)*, 310–339.

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. US Patent 6,090,051.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, (pp. 7.5 – 7.9). New York: IEEE.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics 27*, 89–110.

Rubio-Fernández P. (2016). How Redundant Are Redundant Color Adjectives? An Efficiency-Based Analysis of Color Overspecification. *Frontiers in Psychology 7:153*. doi: 10.3389/fpsyg.2016.00153

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual interpretation. *Cognition, 71*, 109–147.

Shannon, C. E. (1948), A Mathematical Theory of Communication. *Bell System Technical Journal*, 27: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Tarenskeen S., Broersma M., & Geurts B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology 6:1703*. doi: 10.3389/fpsyg.2015.01703

Tourtouri, E. N., Delogu, F., & Crocker, M. W. (2015). ERP indices of situated reference in visual contexts. *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2422-2427). Austin, TX: Cognitive Science Society.