Rational Redundancy in Situated Communication

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes
vorgelegt von
Elli Nantia Tourtouri
aus Drama, Griechenland

Saarbrücken, 2020

Der Dekan: Prof. Dr. Heinrich Schlange-Schöningen
Berichterstatter/innen: Prof. Dr. Matthew W. Crocker, Prof. Dr. Antje S. Meyer
Tag der letzten Prüfungsleistung: 02.07.2020

Abstract

Contrary to the Gricean maxims of Quantity (Grice, 1975), it has been repeatedly shown that speakers often include redundant information in their utterances (over-specifications). Previous research on referential communication has long debated whether this redundancy is the result of speaker-internal or addressee-oriented processes, while it is also unclear whether referential redundancy hinders or facilitates comprehension.

We present a bounded-rational account of referential redundancy, according to which any word in an utterance, even if it is redundant, can be beneficial to comprehension, to the extent that it facilitates the reduction of listeners' uncertainty regarding the target referent in a co-present visual scene. Information-theoretic metrics, such as Shannon's entropy (Shannon, 1948), were employed in order to quantify this uncertainty in bits of information, and gain an estimate of the cognitive effort related to referential processing. Under this account, speakers may, therefore, utilise redundant adjectives in order to reduce the visually-determined entropy (and thereby their listeners' cognitive effort) more uniformly across their utterances.

In a series of experiments, we examined both the comprehension and the production of over-specifications in complex visual contexts. Our findings are in line with the bounded-rational account. Specifically, we present evidence that: (a) in view of complex visual scenes, listeners' processing and identification of the target referent may be facilitated by the use of redundant adjectives, as well as by a more uniform reduction of uncertainty across the utterance, and (b) that, while both speaker-internal and addressee-oriented processes are at play in the production of over-specifications, listeners' processing concerns may also influence the encoding of redundant adjectives, at least for some speakers, who encode redundant adjectives more frequently when these adjectives contribute to a more uniform reduction of referential entropy.

Acknowledgements

I would like to start by thanking my supervisor, Matthew Crocker, first for believing in me and offering me a position in his group, but also for making me part of the SFB project. Matt has been very supportive of my ideas, very generous with offering resources for my research or training, very inventive in finding solutions when I thought there were none – but also with coming up with problems when I thought that everything was solved! I thank him for all!

My endless gratitude goes to my friend, office-mate, and co-author, Francesca Delogu, one of the most generous, kind and hard-working people I know. I was enormously lucky to have met her and worked with her. Thank you, Fra, for being my tireless teacher, for all the times you helped me out with smaller or bigger problems, and searched solutions for me as if it were for yourself. Thank you for listening to me all the times I needed an audience to work out some idea, and all the other times I just needed an audience! I'll miss seeing you every morning, but never forget: Το παθος για τη λευτερια ειναι δυνατοτερο απ′ ολα τα κελια!

Les Sikos, my other co-author, is 'responsible' for my first (actual) production experiments. These were uncharted waters for me, but Les' contribution made the journey easier. So thanks, Les, for your all your ideas and advice, for being so patient every time I appeared in your office with a new puzzle, and for all the proof-reading and editing!

I'm grateful to Mindaugas Mozuraitis, for his help with everything related to eye-tracking and for generously sharing all his wisdom on statistical analyses, and Maria Staudte for her help with the visual-world eye-tracking analysis and lots of useful discussions. Special thanks goes to Harm Brouwer, my go-to math person, always willing to extend a helping hand any time I needed to calculate something or discuss modelling ideas and stuffz (though I'm sure he wouldn't call any of that 'math' or 'modelling')!

I'm also thankful to Mirjana Sekicki and Janis Kalofolias, who were there every time I needed to pick their brains, and offered immense help with the analysis of eye-tracking and ICA data. Many thanks to Vera Demberg, Nikos Engonopoulos, Ben Peters, Jorrig Vogels for demystifying the ICA. My fellow PhD students Yoana Vergilova and Torsten Jachmann, thank you for all the EEG-lab+ troubleshooting sessions ;) Torsten, thanks for taking care of the Zusammenfassung, too!

I'd also like to thank all the brilliant and talented people who were members of the Psycholinguistics Group and our paths crossed at some point along the years: Christoph Aurnhammer, Jesus Calvillo, Heiner Drenhaus, Chiara Gambi, Muqing Li, Ron Pomper, Lisa Teunissen, Elisabeth Rabs, and Noortje Venhuizen. Thank you all for being there with comments and ideas any time I had some results to share or some presentation to practice.

Many thanks also to our Systemgroup team: Christoph Clodo, Bernd Mechenbier, and Jochen Pätzold, who would appear at the doorstep of the lab only seconds after calling for help and do their magic, and to our SFB coordinators Marie-Ann Kühne and Patricia Borrull Enguix, as well as to our secretary Claudia Verburg, for taking care of bureaucratic and administrative hassles and making everything so much easier!

For help with creating the audio stimuli of my experiments, I'd like to thank Steffi Nickels and Andreas Conrad, as well as Jeanin Jügler. For help with creating the visual stimuli and implementing the first EEG experiment I thank – our then students – Yoav Binoun, Torsten Jachmann, and Zijian Lu. I would also like to thank all the HiWis who helped out with collecting, transcribing and annotating data: Julia Kolb, Christine Muljadi, Polina Stadnikova and Daria Stesenko. And thanks to everyone who participated at some version of my experiments!

My dear friends and fellows Stavroula Alexandropoulou and Myrto Pantazi, thank you guys for being there all these years, for proof reading every little thing I'd send you, and for the countless times we discussed this project. You already know everything; feel free to not read it :P

Finally, to my parents, Despina and Spyros, and my sister, Myrto, I owe everything. Thank you for your endless love and support. Thank you for believing in me, for seeing the best in me and for always challenging me to become better!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In a scene from the ’90s TV series *Twin Peaks*, Sheriff Harry Truman walks into the local police station moments after the receptionist, Lucy Moran, has received an alarming phone call. She immediately informs him of the call and says ‘I’m gonna transfer it to the phone on the table by the red chair; the red chair against the wall; the little table with the lamp on it; the lamp that we moved from the corner’. As the Sheriff stands still staring at her, she finally yells ‘The black phone, not the brown phone!’.[1]

What makes this awkward exchange deserve its own video on YouTube? Let us take a closer look. As it becomes clear at the end, Lucy aimed at specifying one out of two phones. A rather plain task. What is interesting, however, is her choice of words. In an initial attempt, she described the intended device using location information (‘the phone on the table by the red chair’). Even though this information was, arguably, enough for the Sheriff to figure out which of the two phones he had to answer, Lucy kept adding details. But to no avail; the Sheriff seemed very uncertain about what he was asked to do. On a second try, Lucy specified the colour of the intended phone (‘the black phone’). Although this information was also sufficient to distinguish between the two phones, she complemented this description, too (‘not the brown phone’). While this is an extreme case – which has, nonetheless, earned this video 17000 views! – we all have found ourselves in similar situations, where we said more than needed as we strived to be understood, raising the question:

[1] For clarity, we use the feminine pronoun to refer to the speaker (Lucy), and the masculine pronoun to refer to the addressee (Harry) throughout.

What motivates such choices? That is, how do speakers select between more or less explicit descriptions, and why do they often say more than necessary?

This thesis is concerned with the use of redundancy in definite referring expressions and seeks to explain why it occurs, especially given that according to traditional pragmatic theories (Grice, 1975) speakers should avoid unnecessary prolixity. In search of an answer, we wonder how referential redundancy influences comprehension processes. We turn, therefore, to both the production and comprehension of redundant referring expressions, and tackle them experimentally. We focus, specifically, on situations where the meaning of a referring expression is determined not only based on the meaning of its words but also based on the visual environment wherein it is uttered. We conclude by offering a bounded-rational account of referential redundancy that is based on Information Theory and unifies comprehension and production processes. According to this account, speakers may include redundant words in their utterances, when these words help to manage the listeners' uncertainty about what is being communicated (i.e., the identity of the target referent).

1.1 Reference Production in Visual Contexts

Most theories of language production agree that speaking is a process that proceeds in roughly three stages – each with its own subprocesses. These stages generally involve the conceptual preparation of an intended message, the encoding of this message into less and less abstract linguistic forms (e.g., morphological, phonological and phonetic encoding), and the realisation of that message – i.e., articulation (e.g., Dell, 1986; Garrett, 1988; Levelt, 1989; Levelt et al., 1999). Under most language production models, each stage generates an output representation, which serves as input into the next stage (see Fig. 1.1). In Levelt's model (Levelt, 1989; Levelt et al., 1999) – the most influential and widely accepted account of speech production (see Meyer et al., 2019) – the intended message is formulated at a first stage, and activates the relevant *lexical concept* (e.g., ESCORT(X,Y)).[2] The lemma linked to this lexical concept is then accessed (e.g., *escort*), and its syntax becomes available. At a next stage, the word form is retrieved, and its morphological and phonological information become available, i.e., the word's morphemes (e.g., <escort>, <ing>) and phonological makeup (e.g., /ə/, /s/, /k/, /ɔ/, /r/, /t/ for <escort>, and /ɪ/, /ŋ/ for <ing>).

[2]Example and notation are following Levelt et al. (1999).

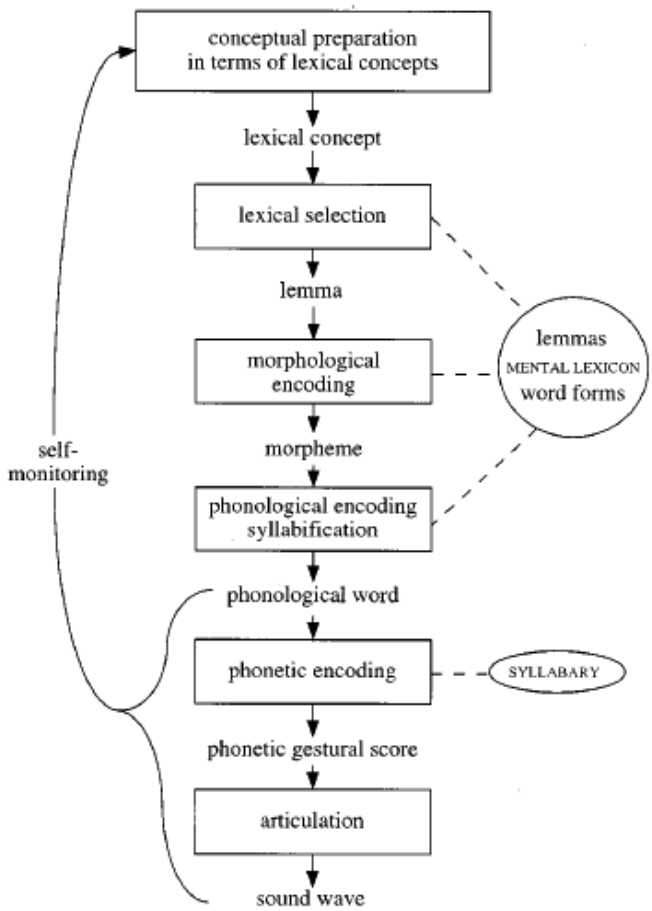


Figure 1.1 Outline of a word production model (Levelt et al., 1999, p. 3).[3]

Finally, a gestural score for the word's syllables (i.e., [ə], [skɔr] and [tiŋ]) is computed (or accessed in the syllabary), and is executed by the articulatory system.

Noticeably, these models centre around the production of single-word utterances. Yet, it is reasonable to assume that the production of multi-word utterances, such as referring expressions, proceeds in a similar fashion. The speaker would still need to go through all of the production stages described above for each word individually, and delay articulation for the first word until the lemma of the upcoming word has been retrieved, in order for the words in the utterance to be temporally coordinated (see Levelt and Meyer, 2000; Meyer, 1996; Schriefers, 1993).

[3]Permission was acquired as necessary for the reprint of this and all subsequent figures from other work.

In any case, referential choice is hypothesised to take place at a very early stage in language production, namely during conceptual preparation (cf. Schmitt et al., 1999). That is, as a first step in producing a referring expression, speakers must choose from within a set of more or less explicit alternatives the one that is more appropriate to refer to the intended entity. If they wish, for example, to refer to a ball, they can do so using the word 'ball', but they could also use the hypernym 'toy', a simple pronoun, like 'it', or a more specific expression, like 'John's blue striped ball'. It is, however, a matter of ongoing debate which factors determine the explicitness of speakers' referring expressions and exactly what processes are involved.

Research on reference production in discourse has demonstrated that factors such as the status of the intended referent affect speakers' choice of referring expression. This research suggests that information that is *prominent* (or accessible, salient, etc.) is more likely to be later referred to with a reduced form like a pronoun (Ariel, 1988, 2001; Brennan, 1995; Gundel et al., 1993). That is, entities that have a more prominent conceptual or psychological status (i.e., are more accessible in discourse or in memory) require less detail in order to be identified (see Arnold, 2008 and Arnold and Zerkle, 2019 for review and discussion of these accounts).

But what about referential choices relative to entities in the visual context? Arguably, the visual environment renders co-present entities equally accessible: they are all right before our eyes. That is, there is no need of maintaining their representations in memory, as is argued to be the case in the discourse reference tradition (see Davies and Arnold, 2019 for a review and comparison of the two traditions). So, what is it that determines how explicit a referring expression should be in visually-situated communication?

1.1.1 Common Ground

Entities in the visual world can also be thought of as more or less prominent, in terms of their *visual* salience or prominence (i.e., the salience due to perceptual properties, such as their size, colour, etc.), which may affect speakers' referential choices. That is, referents that are larger or brighter are visually more prominent, and these properties may, therefore, be more likely to attract attention and be encoded in the referring expression, even redundantly (see Belke and Meyer, 2002; Pechmann, 1989). However, visual salience is not an absolute property of the target referent, but depends on the properties of the other objects in its visual context: a skyscraper would stick out in the skyline of Saarbrücken, but not in New York.

In visually-situated communication, the shared visual context – and the mutual belief that this context is shared – are part of the *common ground* (see also Clarke et al., 2015; Richardson et al., 2009), which is generally defined as the knowledge, beliefs and assumptions shared between interlocutors (Clark and Marshall, 1981). In real-time face-to-face communication, common ground needs to be updated on a moment-by-moment basis, a process known as *grounding* (Clark and Brennan, 1991). According to Clark and colleagues (see also Brennan and Clark, 1996), grounding is a collaborative task, shaped by the purposes of the communicative exchange, that allows interlocutors to establish the mutual belief that the listener has understood the speaker's meaning: The speaker presents the listener with an utterance (presentation phase), under the assumption that the listener will provide evidence that he has understood what the speaker meant with that utterance (acceptance phase).

When the shared task requires interlocutors to *establish reference*, they need to reach the mutual belief that the listener has understood the speaker's message well enough, and is able to identify the intended referent in the visual context. Let us return to the *Twin Peaks* scene from above, and consider as an example how grounding takes place there. At first, Lucy presented the Sheriff with a reference in four instalments, expecting him to confirm that he had understood her utterance and accepted her expression as reference to the particular object in the world after each of them. As he failed to provide such evidence, she acknowledged that grounding failed, and went on to reformulate her message, this time using colour information. This new reference was successfully grounded, as the Sheriff's swift move towards the target phone indicated. What is evident from this minimal interaction is that establishing reference during situated communication is not a task that concerns only the speaker – as is language production in the case of, e.g., talking about one's ideas, giving a speech or writing a book – but it actively involves the listener, too.

Traditional theories of language production (cf. Dell, 1986; Garrett, 1988; Levelt, 1989) have largely ignored the role of the *listener* in utterance planning. All of the production stages outlined above are regarded as internal to the speaker, and the listener is called into action only after articulation, when language comprehension processes commence. Speaking and listening were largely treated as two distinct processes taking place in turns without much interaction. However, communication is, as discussed above, a joint activity, the success of which requires speakers and listeners to collaborate and coordinate (Clark, 1996). Speakers, in particular, need to monitor that they are understood in real time (see Clark and Krych, 2004). Such a monitoring mechanism can, in principle, occur either before (covert monitoring) or after (overt monitoring) articulation in a language production model (see Fig. 1.1).

One way for speakers to ensure that they are understood would be to wait for feedback, after the addressees have processed their utterance. This method is compatible with an overt monitoring mechanism, where speakers can spot any errors while listening to their speech stream, in much the same way they would if they were listening to someone else speaking. Under this view, concerns regarding message interpretation do not influence speakers' referential choices; if they are poorly understood, they will be asked for clarification. Any rough parts can, then, be traced back and modified. Alternatively, speakers can take into account the perspective of their addressees proactively. Given what information is in common ground, they can formulate assumptions about their addressees' knowledge, and based on these assumptions imagine whether the utterance they are about to produce will be well understood. This view is compatible with a covert monitoring mechanism, allowing speakers to track any problems with, e.g., lexical selection, before they make their utterance available to their listener. Information about the addressees' background knowledge and the common ground is, therefore, available early on, and can influence the initial stages of language production (Brennan and Clark, 1996; Clark, 1996; Clark and Marshall, 1981; Galati and Brennan, 2010; Heller et al., 2012).

To summarise, in visually-situated communication the shared visual environment is part of the common ground, i.e., the information and beliefs shared between interlocutors. One question is, then, raised: Does information in common ground and listeners' needs affect speakers' initial production choices or do they only come into play post hoc? In other words, who do speakers plan their utterances for?

1.1.2 Reference Production: For the Speaker or for the Listener?

While it is generally accepted that speakers consider common ground and eventually cater to the needs of their addressees (Arnold, 2008; Horton and Keysar, 1996), there is no consensus regarding (a) how early this process occurs during utterance planning (i.e., whether it is automatic or takes place later on), and (b) what kind of mental representation it is based on (i.e., whether speakers maintain a mental model only for themselves or they also keep track of their addressees' background knowledge). Evidence is to a great extent based on research employing the referential communication task (Krauss and Glucksberg, 1977; Krauss and Weinheimer, 1964). In this task, participants collaborate with each other, in order to manipulate a set of objects or pictures in a workspace or computer screen in front of them. They therefore need to produce – or interpret, as listeners – a series of referring expressions about a restricted set of referential candidates.

Evidence falls under two general views. The first one holds that referential choices are conditioned by a desire to minimise effort for speakers themselves (see Epley et al., 2004; Keysar et al., 1998), and listener needs are not taken into consideration during initial utterance planning, as this would increase demands on the speakers' cognitive resources (e.g., Horton and Gerrig, 2005; Horton and Keysar, 1996; Wardlow Lane and Ferreira, 2008). This *egocentric* account is compatible with the existence of an overt monitoring mechanism. According to the second view, speakers tailor their utterances to the needs of their addressees, a strategy known as *audience design* (Clark & Murphy, 1982). That is, speakers design their referring expressions such that their addressees can efficiently (i.e., quickly, easily and accurately) identify the intended referent. In this view, speakers utilise information that is in common ground in order to formulate assumptions about their addressees' knowledge (Clark, 1996; Clark & Brennan, 1991). Audience design is compatible with covert monitoring.

The egocentric account is supported by evidence showing that speakers' production choices are frequently at odds with listener preferences. Ferreira et al. (2005), for instance, examined speakers' ability to avoid ambiguity when referring to entities in a visual scene. The results showed that while speakers effectively avoided conceptual ambiguity (e.g., were able to distinguish between a small and a large bat), they did not always succeed in avoiding linguistic ambiguity (e.g., distinguishing between a baseball bat and an animal bat), even though they were able to detect the second kind of ambiguity after-the-fact. This behaviour was interpreted as indicating a 'production-based ambiguity detection strategy' (Ferreira et al., 2005, p. 275). In other words, there was evidence that speakers' referential choices were shaped by production-internal constraints, and were not influenced by comprehension-related processes. Using the perspective-taking task,[4] Horton and Keysar (1996) demonstrated that, even though speakers were able to tailor their utterances to the perspective of their addressees, they were less likely to do so under time pressure. This result was taken to suggest that audience-design does not occur during initial utterance planning, but rather takes place at a later stage, where speakers can – if needed – adjust their original expression (see also Epley et al., 2004; Keysar and Horton, 1998; as well as Rubio-Fernández, 2008 for a critical review). Further research has also supported the claim that, even though speakers are capable of audience-design, they are more likely to produce egocentric utterances, especially when they are under cognitive load (Roßnagel, 2000; Vogels et al., 2015), when the addressee's perspective is not deemed relevant to the communicative goal (Yoon et al., 2012), or when the entities in their privileged ground are salient (Wardlow Lane and Ferreira, 2008). These findings, therefore, come in support of a two-stage production model (see Bard and Aylett, 2004; Epley et al., 2004), where audience-design takes place at a later stage. The default production processes are, however, speaker-oriented, and speakers fall back to this default when their cognitive resources are limited or when there is no apparent advantage to audience-design. According to this view, speakers do not need to maintain a separate mental representation for their addressees; if they are later on required to adjust their utterances, they can use their own representation as a proxy to that of their addressees (Brown and Dell, 1987; Fukumura and van Gompel, 2012; Vogels, 2014). Speakers' and listeners' models of the communicative situation are, anyway, thought to be closely aligned (Pickering and Garrod, 2004).

Even though in many cases the availability of information in the speaker's privileged ground may lead to the use of descriptions that are based on the speaker's point of view, (early) production processes need not be egocentric. Other research suggests that taking the addressee's perspective should not necessarily occur as a later adjustment to the speaker's own perspective, but it can take place early on in production (Brennan and Hanna, 2009; Brown-Schmidt and Hanna, 2011; Galati and Brennan, 2010; Heller et al., 2012; Nadig and Sedivy, 2002; Vanlangendonck

[4] In this paradigm, interlocutors view a set of objects arranged in cubbyholes in a workspace or computer screen in front of them. Crucially, not all of the objects are visible to both participants: some objects are shared between both participants (common ground), while others are visible to only one of them (privileged ground). Researchers manipulate whether a contrast between two referents is in common or in privileged ground, and measure whether speakers' references are influenced by their own egocentric perspective or are based on common ground (see Brown-Schmidt and Heller, 2018 for a recent review).

et al., 2016). For instance, using eye-tracking Vanlangendonck et al. (2016) found that speakers were less likely to fixate competitor objects, when these objects were occluded to their addressees compared to when they were visible to both interlocutors. Furthermore, no differences were observed in the duration of utterance planning between the audience-design conditions (where a shift of perspective was obligatory for the expression to be successful) and the control conditions (where no perspective-taking was necessary). In other words, taking the addressee's perspective in designing a referring expression need not be more time-consuming or resource-intensive for the speaker.

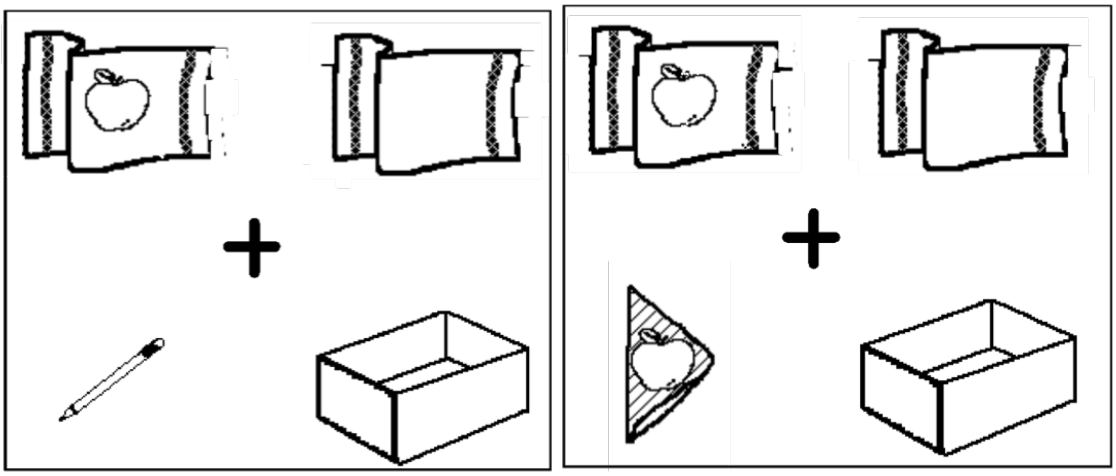
Such flexibility in perspective-taking arguably requires the speaker to maintain two mental representations: one for themselves and one for their addressees (see Clark & Marshall, 1981; Clark & Krych, 2004; Galati & Brennan, 2010). This hypothesis is further supported by evidence that speakers can utilise general or specific information about their addressees in order to draw assumptions about their background knowledge at either a global or a local level (see Arnold, 2008; Davies and Arnold, 2019). That is, speakers use general criteria, such as community membership, in order to draw global assumptions about their addressees' knowledge, and determine the explicitness of their referring expression. For example, New Yorkers were found to use proper names in reference to New York landmarks more frequently when talking to fellow New Yorkers than to non-New Yorkers (Isaacs and Clark, 1987). On the local level, speakers can make use of their prior experience with a specific addressee in order to work out detailed assumptions regarding, e.g., how to refer to previously-mentioned objects. (Clark and Brennan, 1991; Clark and Wilkes-Gibbs, 1986). Additionally, referent identification is found to be more successful when the expression was designed for the specific addressee compared to when it was meant for someone else (Fussell and Krauss, 1989), suggesting that speakers indeed consider their addressees' level of knowledge when constructing a message (see also Schober and Clark, 1989).

In sum, while there is no agreement regarding when (late or early) and how (based on one or two mental representations) perspective-taking occurs during utterance planning, it is generally accepted that speakers do eventually cater to the needs of their addressees. Any account that aims at evaluating audience-design in language production should, therefore, take into account comprehension processes as well. For this reason, we will now present a brief overview of the current knowledge on visually-situated language comprehension, before introducing the problem of referential redundancy in more detail.

1.2 Language Comprehension in Visual Contexts

For a long time, the dominant view regarding real-time language comprehension was language-centric, and did not offer any predictions regarding how non-linguistic visual cues may inform the listener's interpretation (see Knoeferle, 2015, and references therein). However, the development of the *visual-world paradigm* in the mid-'90s (Allopenna et al., 1998; Cooper, 1974; Eberhard et al., 1995; Tanenhaus et al., 1995; see also Huettig et al., 2011 for a review) has allowed researchers to investigate how visually-situated comprehension takes place (see Knoeferle, 2015; Knoeferle and Guerra, 2016). The visual-world paradigm uses eye-tracking technology to record participants' gaze as they scan a scene on a computer monitor or workspace in front of them while listening to scene-related linguistic stimuli. Usually, participants will also be required to perform a task, such as move a specified object from one location in the scene to another. The x and y coordinates as well as the time stamps of participants' fixations are retrieved, and fixations to the object referents are aligned to the onsets of the spoken words. In this way, researchers can gain insights into real-time language comprehension, under the assumption that lexical activation of a name in the linguistic stimulus will cause a shift of overt attention to the relevant referent in the visual scene (Tanenhaus et al., 2000). The most common measure used in visual-world eye-tracking studies is, therefore, the likelihood that a referent will be fixated after the onset of the critical word; referents that match the listeners' on-line interpretation receive more fixations.

Research using the visual-world paradigm has shown that listeners readily use non-linguistic, *visual* features to guide their interpretation of the linguistic input. For instance, when knowledge of the language alone does not allow listeners to arrive at a meaning, as in the case of structural ambiguity, listeners may utilise information in the visual context in order to resolve the ambiguity (Eberhard et al., 1995; Tanenhaus et al., 1995). In one of the first visual-world eye-tracking experiments, for instance, Tanenhaus et al. (1995) had participants listen to sentences such as 'Put the apple on the towel in the box', which at the point of the first prepositional phrase ('on the towel') is ambiguous between a destination and a location reading. That is, before hearing the second prepositional phrase, it was not clear whether the unfolding



(a) Scene depicting single apple

(b) Scene depicting two apples

Figure 1.2 Sample visual scenes paired with the spoken instruction 'Put the apple on the towel in the box' (Tanenhaus et al., 1995, p. 1633).

utterance was asking the listener to move an apple from some unknown location *to a towel* or to move an apple *that was on a towel* to another location. Concurrent with the presentation of the linguistic stimulus, participants viewed visual scenes depicting either a single apple that was on a towel and one empty towel (see Fig. 1.2a), or two apples – one on a towel and the other one on a napkin (see Fig. 1.2b). In both cases, the scene also presented an empty box. Interestingly, as soon as they heard 'on the towel', participants looked more at the empty towel in the one-apple condition, while the were more likely to fixate the apple that was on the towel in the two-apple condition. This result indicates that, in the face of linguistic ambiguity (i.e., when hearing 'the apple' in the context of two apples), listeners rapidly utilised cues from the visual context in order to guide their interpretation of the utterance (i.e., interpret the prepositional phrase 'on the towel' as a modifier to the preceding noun). Similarly, visually depicted actions can affect the incremental assignment of thematic roles, i.e., whether an entity is interpreted as the agent or patient of an action, during ambiguous parts of the sentence (Knoeferle et al., 2005). These studies indicate that the influence of visual information on language processing is immediate.

Other eye-tracking research has demonstrated that visually-situated comprehension is *incremental*. That is, listeners construct their interpretation of the unfolding utterance on the fly, as each incoming word becomes available. In their first visual-world eye-tracking study, Tanenhaus and colleagues directly assessed incrementality in situated comprehension (Eberhard et al., 1995). In one experiment, participants listened to spoken instructions to touch blocks of various shapes arranged in different visual contexts, e.g., 'Touch the starred yellow square'. Depending on the visual context, referent resolution could be *early* (only one starred block was available), *mid* (all blocks were starred, but only one of them was yellow), or *late* (two starred blocks were yellow, but only one of them was a square). The researchers predicted that if the instruction was interpreted incrementally, listeners should not wait to hear the head noun, but should be able to identify the target referent immediately after the disambiguating information was presented. Indeed, participants fixated the target objects shortly after the onset of the disambiguating word in each condition. This finding, therefore, suggests a tight coupling between eye movements to referents in the visual context and the words that denote those referents, such that visual information is rapidly integrated with the linguistic input as it unfolds over time.[5]

The integration of linguistic and visual information has been found to occur at a fine degree of granularity, even at the level of syllables. Allopenna et al. (1998), for example, found that, when listening to the word 'beaker', participants inspected both the picture of a beaker and that of a phonological neighbour ('beetle') more frequently than the pictures of a rhyme competitor ('speaker') or those of unrelated objects. This shift of visual attention started at around 200 ms after the onset of the target word and lasted approximately until word offset. That is, during the first few hundred milliseconds and until the syllabic context disambiguated the target referent, the first syllable was ambiguous between two referential candidates, and listeners considered both of them (and only them) as potential targets. This result suggests that the influence of the visual context on language comprehension processes is early and extremely rapid.

Further studies have shown that the interaction between language processing and visual perception is *bidirectional*. That is, as the visual context guides incremental language interpretation, real-time language comprehension can also influence visual

[5]It should be taken into account, however, that it takes approximately 200 ms to plan and execute an eye movement (Matin et al., 1993).

perception (see Reali et al., 2006; Spivey et al., 2001). Research on visual perception has shown that when a target referent is defined using a conjunction of features (e.g., 'red and vertical') search time is longer compared to when it is defined by a single feature (e.g., 'red'), while it is also affected by the number of distractor objects in the visual scene (see Treisman and Gelade, 1980; Wolfe, 1994, 1998). In visual perception research, the target object is usually described before the start of visual search, i.e., the presentation of the linguistic and visual stimuli is not simultaneous. It has been shown, however, that when the object description is concurrent with the visual search viewers are faster to identify the target object. Spivey et al. (2001) asked participants to look for a '*red vertical*' in a visual context containing red and green bars that were either horizontal or vertical. The researchers manipulated whether the audio instruction preceded the visual display or whether they were concurrent. An additional manipulation concerned the referential set size, which could contain either 5 or 20 bars. Interestingly, when the linguistic and visual input were presented concurrently, visual search was less affected by the number of distractor objects. This result indicates that search for the target referent can begin once the first feature is presented, and proceed in parallel with the presentation of the second feature. Crucially, after the second feature is presented, search is carried out within a subset of the initial set of referents, since the referents that do not match with the first feature have been excluded.[6] These findings, therefore, suggest that linguistic processing and visual perception rapidly interact and one can constrain the other.

Visual-world eye-tracking research has also contributed to the growing body of evidence showing that language comprehension processes are *anticipatory* (DeLong et al., 2005; Van Berkum et al., 2005; see also Hauk, 2005). Altmann and Kamide (1999) showed, for instance, that listeners can utilise verb selectional restrictions in order to predict upcoming content. In visual scenes depicting a boy and four objects, one of which edible (e.g., a cake), after hearing 'The boy will eat...', listeners were more likely to fixate the edible object compared to the others. By contrast, if the verb was non-restrictive (e.g., move), listeners' attention was distributed evenly across all objects. Further research has shown that listeners can also combine word meanings in order to restrict the interpretation of the unfolding sentence and predict upcoming material. In a study by Kamide et al. (2003), for instance, when hearing the verb 'ride' participants looked more frequently at a *motorbike* if the verb subject was '*the man*', but they were more likely to fixate a *carousel* if the subject was '*the girl*'. Participants were, therefore, able to predict which of the depicted 'ridable'

[6] See, however, Gibson et al. (2005) where these results were not replicated with faster speech rates.

objects would be mentioned next drawing on the combined information of the verb and subject, and their knowledge of the world (i.e., men ride bikes, and girls ride carousels – not the other way around). These results suggest that visual information is not only used early and rapidly in language comprehension, but that it can also restrict the unfolding linguistic interpretation and lead to the generation of specific expectations regarding possible continuations of the utterance.

In sum, visual-world eye-tracking studies have provided a wealth of evidence regarding real-time language processing. First, (visually-situated) language comprehension is highly incremental, occurring even at the level of phonemes. Secondly, the link between the linguistic and visual streams is bidirectional: linguistic cues are used to guide the listeners' visual search for the intended referent, but also cues from the visual environment are utilised to inform the interpretation of the unfolding utterance. Lastly, there is robust evidence that language processing is predictive, and that listeners can combine information in the linguistic and visual context to predict upcoming linguistic material.

1.3 Is it Irrational to be Redundant?

Let us now return to the issue of referential redundancy. Given the current knowledge on production and comprehension processes, why are overly explicit referring expressions considered problematic? The *context* in which communication takes place is key to answering this question. As we saw above, the visual context is part of the speakers' and listeners' common ground, affecting both speakers' referential choices and listeners' on-line comprehension. The context is, therefore, thought to set the level of explicitness required to refer to a specific object, with more explicit referring expressions serving the purpose of distinguishing between similar referents (see Olson, 1970). For example, in a visual context containing a single ball, mentioning any feature other than its type would be unnecessary to refer to this object: an expression such as 'the blue ball' would be *over-specified*. However, if the context also contained a green ball, '*blue*' would be necessary, and the same expression would in this case be *minimally-specified*.

Traditional (pragmatic) accounts of communication support the idea that if speakers wish to get a message across, they should choose encodings that are appropriately specific (see among others Altmann and Steedman, 1988; Crain and Steedman, 1985;

Grice, 1975; Horn, 1984, 2004; Levinson, 1983, 2000). These accounts largely resonate Grice's influential theory of conversational implicatures (Grice, 1989, 1975), laid out in more detail in Section 2.1. In particular with regard to redundancy, Grice's theory postulates that, for communication to be successful speakers should convey the *minimal* amount of information that is necessary. The premise is that people engage in conversation bearing in mind a set of conversational 'best practices', to which they conform and *expect* their partner to also conform to. That is, if the speaker defies one of these principles, the listener will reason about it (whether or not the speaker could observe the principle, but chose not to), which may lead them to infer some implicit but intended meaning. By this logic, when the speaker uses a referring expression that is more specific than warranted by context, listeners may experience difficulty with understanding: As listeners make constant use of visual cues around them in order to inform their incremental interpretation, they can readily realise that not all information in the expression was required, nor was there some implicit meaning to be recovered.

But, this is in theory. In practice, do speakers and listeners live by principles that specify how much information is acceptable to be conveyed in an utterance? There is growing evidence that speakers are not Gricean, as they frequently produce expressions that are more explicit than necessary, i.e., they over-specify (e.g., Deutsch and Pechmann, 1982; Engelhardt et al., 2006; Koolen et al., 2011; Pechmann, 1989; Rubio-Fernández, 2016; Tarenskeen et al., 2015). It is, however, not clear whether listeners' comprehension is affected by such over-specifications: Some research shows that listeners may be confused by the use of redundant information and their comprehension may, thus, be impeded (e.g., Davies and Katsos, 2013; Engelhardt et al., 2011), while other work finds that over-specifications do not actually have a negative impact on comprehension (e.g., Arts et al., 2011a; Brodbeck et al., 2015). Understanding how over-specifications influence comprehension is important for understanding why they occur. If comprehension is hindered, it may be argued that over-specifications are due to production-internal purposes; if comprehension is facilitated, over-specifications may be attributed to an audience-design strategy (see Davies and Arnold, 2019).

Human behaviour is generally thought to be governed by a desire to expend as little effort as possible in order to achieve one's goals (*principle of least effort*, Zipf, 1949). In linguistic interaction, the speakers' objective would, therefore, be to use as few words as possible for the listener to understand their communicative intention. As

Piantadosi et al. (2011) schematically put it, speakers would ideally produce a single, maximally ambiguous word expressing every possible meaning. It is, therefore, not clear why speakers would choose to utter words that are not required, thus increasing their own production effort.

In Grice's view 'talking [is] a special case or variety of purposive, indeed rational, behavior' (Grice, 1975, p. 47), with parallels in domains of human interaction other than verbal communication. In other words, Gricean speakers and listeners are rational agents who reason about their goals, and choose the paths of action that will allow them to achieve these goals. Communication in Grice's theory is, therefore, perceived as a *rational activity* (Grandy and Warner, 2017). According to Anderson (1990, 1991), humans are generally rational, because their cognitive system has been evolved so as to optimise behaviour in order to achieve specific goals. This means that in order to solve any given problem people will choose a course of action that approximates the optimal solution. In the case of referential communication, therefore, rational speakers should choose optimal referring expressions conveying *precisely* the information that is relevant for identifying a target referent – as is predicted by the Gricean theory.

Recently, the Rational Speech Acts (RSA) framework (Frank and Goodman, 2012; Goodman and Frank, 2016) provided a formalisation of the Gricean theory. The RSA views speakers as rational agents, who plan their utterances based on a simplified, *literal listener* model. On the comprehension side, listeners use Bayesian inferencing in order to recover speakers' meaning (i.e., the intended referent), based on the particular utterance and the context wherein it was produced. The RSA model for the listener is, therefore, recursive; pragmatic listeners reason about speakers who reason about listeners.

The optimal solution might, however, be elusive, due to various reasons, such as time limitations, the unavailability of crucial information, or because it requires highly complex calculations, etc. (*bounded rationality*, Simon, 1955; see also Chase et al., 1998; Gigerenzer, 1997; Gigerenzer and Selten, 2001). Rational agents may, therefore, resort to the most appropriate solution given current circumstances. A bounded-rational approach may, therefore, be better suited to explain the use of over-specified expressions: Speakers may use redundant information because they lack the time, information or cognitive resources to perform the calculations required to produce a minimal description. At the same time, it is possible that cooperative speakers design their expressions for bounded-rational conversational partners.

Under recent information-theoretic accounts of communication, the cognitive effort that is necessary to process a word in a sentence is associated with the informativity – or 'information load' (see Shannon, 1948) – of that word (Hale, 2003; Levy, 2008). Speakers' production choices may, therefore, be shaped by a concern to manage information load in their utterances, in order to facilitate listeners' comprehension (see Jaeger, 2010; Levy and Jaeger, 2007). As a consequence, over-specifications may be the result of speakers' attempt to distribute or 'stretch out' information load (and processing effort) across the utterance.

What is in such a strategy for the speaker? Clark and colleagues (e.g., Clark and Brennan, 1991; Clark and Krych, 2004; Clark and Wilkes-Gibbs, 1986) propose that by accommodating the addressees' needs, speakers minimise joint effort. That is, production choices are motivated by a desire to minimise the total effort spent for successful communication, in terms of resources, time, etc. By planning their utterances with the listener in mind, speakers benefit themselves, too: they will not be required to repeat the utterance, reformulate the message, or provide more information later on. That is, it is rather a principle of *least joint effort* that speakers abide to, as such a principle serves the purpose of the communicative interaction, and through it, the speakers' own interests.

In sum, given that speakers put in extra effort to produce redundant words quite frequently, the question emerges: Are redundant speakers simply 'irrational', or do they follow some other, bounded-rational, strategy? In this thesis, we set out to investigate the use of referential redundancy using information-theoretic tools in order to determine whether it is a (bounded-)rational behaviour or not.

1.4 Research programme

The main goal of this thesis is to investigate why and when speakers over-specify their referring expressions in visually-situated communication. More specifically, it explores the hypothesis that in view of sufficiently complex visual scenes speakers may follow an audience-design strategy, whereby they utilise redundant words in order to manage the information load associated with processing their referring expressions and with mapping these expressions onto real-world entities.

As detailed above, answering any question about production strategies is essentially entangled with answering how these strategies affect comprehension processes.

Testing our main hypothesis, therefore, rests upon tackling three sequential research questions (RQ), each building on the answer to the previous one.

Because there is no conclusive evidence regarding how over-specified expressions are processed by listeners, our first research question concerns exactly this:

RQ-1 Are over-specifications detrimental or beneficial to comprehension?

In case listeners' comprehension is hindered by the use of redundant information, it may be argued that over-specifications result from an intention to decrease cognitive effort for speakers themselves. If, on the other hand, comprehension is not hindered by redundancy, this could suggest that speakers over-specify in order to facilitate listeners' processing and identification of the intended referent.

Before turning to the production of over-specifications, one more question is necessary. Our hypothesis suggests that speakers' use of redundancy in situated communication should depend on the distributional properties of the visual scene; i.e., speakers should over-specify for features that better manage the information load for the listeners. We, therefore, need to make sure that the way information load is managed through the expression influences participants' processing. Our second research question, therefore, is:

RQ-2 Is the comprehension of referring expressions influenced by how efficiently they manage the visually-determined information load?

If referential processing is found to be affected by the distributional properties of the visual scene, we can then proceed to investigate whether what motivates speakers' use of redundancy is indeed a concern to manage this visually-induced information load more efficiently for their listeners. The third research question we are going to address is:

RQ-3 Do speakers use referential redundancy as a means to modulate the visually-determined information load for their listeners?

If speakers are found to vary their use of redundant utterances based on the structure of the visual context and on how well the redundant word lets them manage the visually-induced information load, this would be evidence for an audience-design production strategy. If, on the other hand, the use of redundancy does not differ with visual scene structure, speakers' over-specifications may be the result of an egocentric strategy.

In order to gain insights into the underlying mechanisms of referential processing (RQ-1 and RQ-2), we draw upon three different dependent measures. In particular, we recorded participants' *event-related brain potentials (ERPs)* to measure word expectancy; *eye movements* as a measure of the differential expectancy for any of the visually co-present referents to be mentioned; and a pupillary measure, the *Index of Cognitive Activity (ICA)*, as a measure of overall cognitive effort.[7] We combined the evidence offered by these three measures regarding the on-line comprehension of over-specifications with production measures, in order to study whether and how comprehension preferences might influence speakers' production choices (RQ-3).

1.5 Thesis overview

The remainder of this thesis will proceed as follows. We will start with reviewing the traditional pragmatic approach to communication, focusing particularly on referential redundancy (Section 2.1). Sections 2.1.1 and 2.1.2 will offer an overview of previous experimental findings on the production and comprehension, respectively, of referential over-specifications. We will, then, briefly outline Information Theory (Section 2.2), and the use of information-theoretic measures, such as surprisal and entropy reduction, in the (computational) psycholinguistic research (Section 2.2.1), as well as the Uniform Information Density hypothesis – an influential approach linking production and comprehension processes (Section 2.2.2). Section 2.3 will then seek to extend these information-theoretic notions into visually-situated settings, and formulate in more detail our hypothesis that over-specification can be accounted for by a bounded-rational approach.

In Chapter 3 we present an initial ERP experiment aiming at identifying the neural correlates of reference resolution, and assessing whether over-specifications in particular are detrimental or beneficial to language comprehension. The results of this experiment will be discussed in the context of different pragmatic and bounded-rational approaches.

Chapter 4 presents two additional comprehension experiments, one eye-tracking and one ERP experiment. These experiments address how referential specificity and visually-determined information load independently influence on-line com-

[7]The precise motivation for the use of each of these measures is detailed in the relevant chapters.

prehension, and whether these two factors interact. The results from these two experiments offer additional methodological insights regarding the comparability of results acquired by different methodologies, and dependent measures.

In Chapter 5 we move to the production side of referential communication, and present two experiments examining whether speakers are sensitive to the distributional properties of the visual scene, and over-specify when the redundant words allow them to manage visually-induced information load for their listeners. In the first experiment, only the head noun is required to identify the referent, while in the second experiment the mention of an attribute is also necessary. Results of both experiments underline the high variability of production choices among speakers, and the second experiment in particular identifies three different production strategies.

In Chapter 6 we first present an overview of our current experimental findings (Section 6.1), and summarise the key contributions of our research, theoretical (Section 6.2.1) and methodological (Section 6.2.2). We then proceed to illustrate how our findings can be brought together with seemingly incompatible results from previous research under a unified account of referential communication (Section 6.3.1). We conclude by offering a bounded-rational account of referential redundancy, and discuss how such an account fits with current research in the field (Section 6.3.2).

Chapter 2

(Rational) Redundancy in Referential Communication

As we saw in Chapter 1, an important aspect of everyday communication is establishing reference, either to entities mentioned in previous discourse, or to objects in our immediate visual context. A key ingredient to successfully establishing reference is common ground, that is the knowledge and beliefs that are shared between speakers and listeners. Egocentric language production accounts argue that speakers' referential choices are not conditioned on common ground or addressees' needs, but are oriented towards facilitating production processes for themselves. Audience-design accounts, on the other hand, hold that speakers consider common ground and tailor their utterances to the needs of their addressees. In visually-situated communication, which is the focus of this thesis, the shared visual context is part of the common ground. In such settings, speakers frequently provide more information than necessary to identify a referent given the context – i.e., they over-specify. This behaviour is problematic for theories of communication, in part because it is not clear whether this redundancy causes comprehension difficulties for the addressees, and raises the question why speakers over-specify. In this Chapter, we briefly review the Gricean pragmatic account, as well as more recent probabilistic accounts of communication, and seek to understand how they treat referential redundancy. In the final section, we seek to extend these probabilistic approaches into visually-situated communication, in order to account for the use of over-specifications.

2.1 The Gricean pragmatic account

In his seminal work on conversational implicature, philosopher Paul Grice postulates that cooperation between speakers and listeners is a prerequisite for successful communication (Grice, 1975; 1989). This idea is expressed in a general cooperative principle, according to which, interlocutors should 'make [their] conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange at which [they] are engaged in' (Grice, 1975, p. 45). The cooperative principle is further specified by four maxims, each pertaining to a different aspect of communicative interaction: Quantity, Quality, Relation and Manner. These maxims require speakers to be informative, truthful, relevant and perspicuous, respectively.

Grice used the example of two people collaborating in mending a car, in order to ground his maxims into analogies with other instances of real-life interactions. If Lucy is assisting Harry to mend a car, Harry expects her to hand him the exact number of screws necessary – no more or less (Quantity); he expects her to hand him screws that are appropriate – and not e.g., bent (Quality); he expects a hammer and not a tennis racquet (Relation); and he expects Lucy to be efficient in completing these tasks (Manner). By analogy, listeners expect speakers to not say more or less than necessary, to deliver information that they know to be true and that is relevant to the goal of their interaction, and finally to do so in an orderly and efficient fashion.

Grice did not intend to devise a theory of how communication takes place, but rather aimed at explaining how conversational *implicatures* are generated. That is, how addressees recover a meaning that was intended by the speaker, but not explicitly expressed in the utterance. For example, if at the dinner table Lucy asks Harry 'Can you please pass the salt?', she expects Harry to recognise that by uttering this question she requests that he pass her the salt; she is not actually interested in learning about his ability to pass the salt. Harry's response should, therefore, be to perform the physical act of reaching for the salt holder and handing it over to Lucy, instead of replying to her question with a 'yes'. Such cases, where the intended meaning needs to be *inferred* by the listener, are called implicatures.

In Grice's theory, implicatures arise when the conversational maxims are *violated*, e.g., due to a clash between different maxims. In this case, the listener will still assume that the speaker is observing the cooperative principle – unless he is convinced for the opposite – and try to work out the speaker's intended meaning. Take for example the following interaction: If Lucy tells Harry 'I have to work tonight', the meaning of her utterance will depend not only on the semantics of the words that she used, but also on the context where it was uttered, i.e., the question that it replies to. If Harry had asked 'What are your plans for tonight?', he is entitled to interpret Lucy's utterance as simply conveying the meaning that she was planning to get down to work. If, however, Harry's question was 'Are you coming to John's party tonight?', Lucy's response would have a different meaning; in this case, Harry can safely infer that Lucy meant to say that she was not going to the party. Even though it was not *literally* expressed in her utterance, Harry can recover this meaning just by knowing the semantics of the sentence and assuming that Lucy is being cooperative; i.e., that she is observing the cooperative principle and the maxim of Relation that follows from it, which requires contributions to be relevant. Therefore, by residing on the assumption that Lucy is cooperative and on world knowledge that partying and working do not go together, Harry is entitled to infer that by saying that she has to work, Lucy *implicates* that she is not going to the party.

Most relevant to the issue of referential redundancy is the category of *Quantity* and its two maxims, according to which, a conversational contribution should be as informative as is required (first maxim), and not more informative than is required (second maxim) for the purposes of the exchange. A violation of the first maxim of Quantity, would leave the listener with less information, and perhaps lead to the generation of a conversational implicature – e.g., that the speaker is being uninformative, thereby violating the maxim of Quantity, because she cannot attest to the accuracy of any additional information, which would cause her to infringe another maxim, namely that of Quality. When communication concerns a referent in the immediate visual context, the listener would expect to receive the amount of information that is required for the identification of that referent. For example, if the speaker intends to specify a blue ball in a context containing a blue and a green ball, the utterance 'Find the ball' would violate the first maxim of Quantity, because in this context, it is *under-specified* – i.e., it carries less information than necessary. According to the Gricean theory, the listener will try to reason about this lack of information. If there is no evidence that his partner is in fact uncooperative, he may draw the conclusion that, for instance, only one of the two balls is visible to her.

It is not clear, on the other hand, what a violation of the second Maxim of Quantity would implicate. Grice himself pointed out that such a violation might not in fact constitute an infringement of the cooperative principle, but may be 'merely a waste of time' (Grice, 1975, p. 46), and that in any case unnecessary prolixity could be avoided based on the maxim of Relation.[1] He recognised, however, that the use of extra information might leave listeners confused about the purpose of this redundancy. For example, if the utterance 'Find the blue ball' was used in the context of a single ball, the adjective would be unnecessary, and the expression would be *over-specified*, violating the second maxim of Quantity. In this case, the listener might be left wondering why the adjective ('blue') was used even though the head noun ('ball') would have sufficed, which could lead to the inference that he does not share the same visual context with the speaker; i.e., that in the speaker's perspective there is a second ball of a different colour. In contrast to cases of under-specification, however, the listener would be able to establish reference, because, despite the redundancy, the information required for target identification *is* encoded in the utterance.

Even though the Gricean theory does not make any predictions regarding the online processing of utterances that violate the maxims (see, however, Noveck and Reboul, 2008), it does have implications for the addressees, in that they should expect speakers to observe the conversational principle and the maxims that follow from it (Grice, 1989, 1975). Redundant information may therefore engage addressees in *unintended pragmatic inferencing* (e.g., that a second ball is present but not visible to them, in the previous example), which will ultimately need to be *cancelled*. However, because language processing is rapid and incremental, it is possible that such pragmatic inferences are generated in real time, and may lead to comprehension difficulties (cf. Sedivy et al., 1999). As we will see in Section 2.1.2, however, whether referential redundancy impedes comprehension or not is open to debate, as empirical studies have to date provided mixed evidence.

In sum, based on (a strict interpretation of) the Gricean account, cooperative speakers should be succinct, and say no more than necessary. When speakers do not conform to this expectation, listeners may pick up on this behaviour and reason about it, which may lead them to derive some implicit meaning. In visually-situated communication, it is generally thought that Gricean listeners should expect speakers' utterances to contain only the information that is necessary to identify a target entity. In such contexts, redundancy may generate unintended pragmatic inferences, which may impede processing for the listeners. In other words, Gricean speakers should produce minimally specifying referring expressions, and Gricean listeners should experience comprehension difficulties when this is not the case.

[1] See Horn (1984, 2004); Levinson (1983, 2000), but also Sperber and Wilson (1986).

2.1.1 The Production of Redundant Reference

Since Grice put forth his cooperative principle, much work has investigated whether speakers do in fact observe the conversational maxims in everyday language use (see among others, Arts et al., 2011b; Belke and Meyer, 2002; Davies and Katsos, 2013; Deutsch and Pechmann, 1982; Engelhardt et al., 2006; Koolen et al., 2011, 2013, 2016; Maes et al., 2004; Pechmann, 1989; Rubio-Fernández, 2016, 2019; Tarenskeen et al., 2015; van Gompel et al., 2019; Vogels et al., 2019; Vonk et al., 1992). Despite the different visual settings, tasks or languages employed, these studies share a common finding: that speakers frequently use redundant information in their referring expressions. This redundancy is not in line with (a strict interpretation of) the Gricean maxims (see, however, Bach, 2006, but also Geurts and Rubio-Fernández, 2015), especially when compared to the low proportion of under-specifications. The consistency with which over-specifications appear in referential communication thus raises the question: Why do speakers over-specify?

Generally speaking, two kinds of explanations have been offered, in line with the accounts of reference production reviewed above (see Section 1.1.2): that over-specifications are the result of production-internal processes (egocentric view) or that they are addressee-oriented (audience-design view). Under the egocentric view, in the presence of a visual display that contains referents differing in various attributes, speakers may start to speak before they have fully scanned the display for possible competitors to their intended referent; they may therefore include attributes that turn out to be unnecessary (cf. Pechmann, 1989). It is also possible that in the interest of easing attribute selection and production processes, speakers use features that are visually salient, such as colour (cf. Belke and Meyer, 2002; Koolen et al., 2016, among others). By contrast, the audience-design account holds that speakers include redundant information in an effort to facilitate comprehension for their addressees, for instance by including properties that allow the addressees to create a mental image of the target referent to guide their visual search (cf. Arts et al., 2011a; Paraboni et al., 2007).

To determine the extent to which egocentric or audience-design concerns underlie referential over-specification, past research has tried to identify which factors contribute to the use of redundancy. That is, if speakers are found to over-specify more frequently when the experimentally manipulated factors are associated with the addressees' performance, this should constitute evidence for the audience-design view. Two studies manipulated exactly this factor between participants. In one experiment, one group of participants was told that they were involved in a long-distance surgery, and were asked to produce referring expressions for a listener performing the operation, while a second group was not given a cover story (Arts et al., 2011b). In another experiment, speakers had to produce written instructions on how to set up an alarm clock for addressees in either of two groups: reading the instructions to learn them, or reading while setting up to clock (Maes et al., 2004). In both cases, speakers who thought that their addressees had to carry out an important and demanding task (i.e., carry out surgery, read-to-learn) were more likely to produce over-specifications. These findings suggest that speakers actually consider their partners' performance in the task, and show that the may produce over-specifications in an attempt to be more clear and assist their addressees.

On the other hand, if the use of over-specification is influenced by factors that are mostly relevant to the speaker, this would provide evidence for the egocentric view. Previous work has also manipulated such factors. For instance, properties of the target object such as cardinality (Koolen et al., 2011), or perceptual features such as colour salience (Belke, 2006; Belke and Meyer, 2002; Tarenskeen et al., 2015) have been shown to affect the rate of over-specifications produced. Other research underlines the role of availability in the production of redundant adjectives, such that properties that are conceptually more available to the speaker (e.g., colour or category) tend to be redundantly included in object descriptions more frequently (e.g., Schriefers and Pechmann, 1988).

As speakers need to contrast the intended referent with the distractor objects in order to identify the properties in which they differ, the role of the visual context in the production of over-specified reference has also been investigated. For instance, some studies (Gatt et al., 2017; Rubio-Fernández, 2016, 2019) have found that the rate of over-specification increased with context size (number of distractors). Furthermore, scene variation was also shown to play a role, as it was more likely for speakers to produce redundant colour adjectives in polychrome compared to monochrome displays (Rubio-Fernández, 2016, 2019), and in displays where more distinguishing properties were relevant for the disambiguation of the target referent (Koolen et al., 2013). Finally, the presence of visual clutter (thematically related objects) was also shown to contribute to the production of redundant references (Koolen et al., 2016). However, such factors are related to perceptual characteristics of the referents, and it is possible that while they ease attribute selection for the speaker, they may also facilitate visual search and target identification for the listener.

In sum, despite Gricean considerations speakers frequently over-specify in referential production studies, but there is no consensus regarding whether the use of redundancy is driven by egocentric or audience-design considerations. We now turn to the question how over-specifications affect comprehension processes.

2.1.2 The Comprehension of Redundant Reference

As mentioned above, existing research is divided over whether referential redundancy impedes comprehension or not. Some studies report that over-specification hinders listeners' online processing and results in slower and less accurate identification of the target referent (cf. Davies and Katsos, 2013; Engelhardt et al., 2011), while other work suggests that over-specification may even facilitate comprehension (cf. Arts et al., 2011a; Brodbeck et al., 2015).

For instance, in an ERP study, Engelhardt et al. (2011) found that when visual scenes contained two objects of different shapes (see Fig. 2.1a), redundant prenominal (colour and size) adjectives yielded larger N400-like amplitudes time-locked to the onset of the adjective when compared to scenes with two objects of the same shape (i.e., where the adjective was required for identifying the target, see Fig. 2.1b). The N400 component is a negative deflection of the EEG signal peaking around 400 ms after the onset of a critical word, and it is generally thought to reflect the degree to which the context supports semantic processing, while larger N400 amplitudes are associated with increased processing difficulty (see Kutas and Federmeier, 2011, for a review). Therefore, Engelhardt and colleagues interpreted this N400-like effect as evidence that over-specifications hinder comprehension. The observation of this effect may, however, hinge on the simplicity of the visual context. Namely, it is possible that extra information was strikingly redundant with visual contexts as highly simplified as the ones used in this experiment (only two objects, differing in a maximum of two features). Moreover, any benefits of over-specification might also emerge on the following noun region, while Engelhardt and colleagues only focused on the adjective.

In a similar vein, Davies and Katsos (2013) found evidence that over-specification was dispreferred by listeners as indicated by the lower ratings and longer response times

(a) Different-shape display - Expression is over-specified

(b) Same-shape display - Expression is minimally-specified

Figure 2.1 Sample visual scenes paired with the expression 'Look to the red star' (Engelhardt et al., 2011, p. 306).

for over-specified compared to minimally-specified utterances. Redundancy also negatively affected response times, with over-specified descriptions taking longer to respond to. Material in this study, however, comprised expressions containing different kinds of adjectives, among which evaluative adjectives like 'modern' or 'unsliced', as well as size adjectives, which are known to invoke a comparison between referents, and are, therefore, more likely to be interpreted contrastively (cf. Sedivy et al., 1999; Sedivy, 2003, 2005).

Other offline and online experiments offer evidence in the opposite direction, namely that over-specification facilitates comprehension. Arts et al. (2011a), for instance, showed that referential redundancy might, actually, be beneficial to understanding, and ease participants' identification of the target referent. In this study, participants viewed displays of four objects differing in colour, size, shape, as well as their position on the vertical and horizontal axis. Their task was to identify the location of a target referent based on a description that they read on the previous display. Identification times were similar in minimally-specified and over-specified descriptions, while in some cases they were even faster for over-specifications. These findings, however, came from an offline measure; i.e., identification times were recorded only after participants were exposed to the linguistic stimulus. Thus, no conclusions about the online processing of the referring expression can be drawn.

In sum, there is conflicting evidence regarding the comprehension of over-specifications, with some studies suggesting that redundancy hinders comprehension and others indicating a facilitation. This evidence, however, comes from experiments that vary in the size of the referent set, and adjectives used. Each of these factors may have contributed to the observed effects.

2.2 Information Theory

Formal approaches to communication, such as Information Theory, have recently been used in psycholinguistics (cf. Hale, 2001, 2003; Jaeger, 2010; Levy, 2008; Piantadosi et al., 2011) as they provide a mathematical framework to characterise information and redundancy in language use. Information Theory was originally proposed by mathematician Claude Shannon as a framework for the study of information transmission in telecommunications (Shannon, 1948). The central problem that Information Theory addresses is how transmitted messages can be reconstructed at the receiver's end with as low an error rate as possible, even if there is added noise in the channel through which they are transmitted (see Figure 2.2). Shannon (1948) postulated that in every communication system there is a maximum quantity of information over time that can be transmitted through the channel with an arbitrarily small error probability. This optimal rate of information transmission, dubbed *channel capacity*, must, therefore, be taken into account when encoding a message: Efficient codes should keep information rate at or close to channel capacity, so that messages can be transmitted with minimal information loss and receivers can reconstruct them accurately.

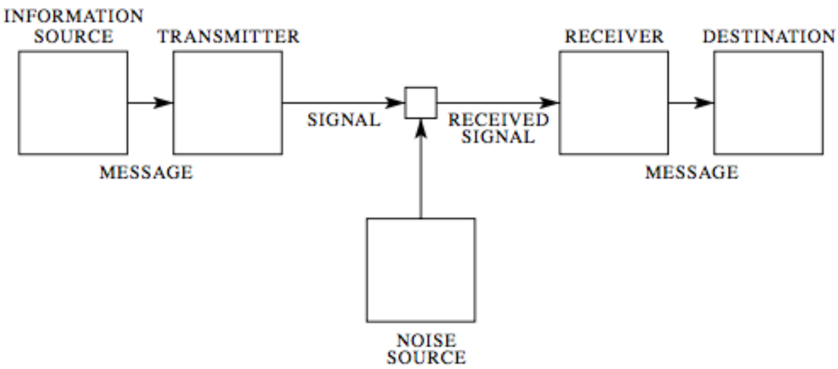


Figure 2.2 Diagram of a communication system (Shannon, 1948, p. 2).

A key notion in Information Theory is that of *entropy*, which offers a quantification of the uncertainty in a system. Entropy is measured in *bits* of information, which can intuitively be thought of as the (average) number of yes-no questions required to determine the state of the system. For instance, the outcome of flipping a coin can be determined by answering only one question (e.g., 'Is it heads?'). Entropy is formally calculated as the negative logarithm of the probability distribution over all potential instances of that outcome, as given in Equation (2.1),

$$H(X) = -\sum_{x \in X} p(x) log_2 p(x) \tag{2.1}$$

where $p(x)$ is the probability of occurrence of event x. Because of the properties of logarithms, Equation (2.1) can be re-written as Equation (2.2),

$$H(X) = \sum_{x \in X} p(x) log_2 \frac{1}{p(x)} \tag{2.2}$$

capturing the fact that the probability of each individual event contributes to the calculation of the total amount of information in the entire distribution. In the coin toss example, for instance, 1 bit of information conveys the outcome of flipping a fair coin. Now, if we had two coins, we would need on average:

$$H(X) = \sum \frac{1}{4} log_2 \frac{1}{\frac{1}{4}} = log_2 4 = 2 \tag{2.3}$$

bits of information to specify the outcome of flipping all of them.

Recent years have seen a revival of Information Theory in the language sciences. Information-theoretic metrics have been used to estimate the *informativity* (information load) of linguistic events – be it phonemes, words or utterances – in terms of their probability to occur in a specific context (see Crocker et al., 2016). For instance, the notion of entropy is used to quantify the uncertainty that comprehenders experience about possible sentence completions at any given point during incremental interpretation, with entropy being higher the more alternatives there are. These metrics have been associated with the cognitive effort required for processing the relative linguistic units, and have served as linking hypotheses connecting comprehension theories to observed data, such as ERP components or reading time data (Hale, 2001, 2003; Levy, 2008).

However, the application of such information-theoretic notions on language processing has so far been language-centric, i.e., it is mostly focused on processing in linguistic contexts. In this thesis, we aim at extending these information-theoretic notions to cases of visually-situated processing, in order to measure the informativity of a given word based on its *visual context*. In what follows, we briefly review recent accounts of language processing that use information-theoretic metrics of cognitive effort, as well as an influential hypothesis linking comprehension and production processes. Lastly, we present an attempt at extending these information-theoretic metrics to visually-situated communication.

2.2.1 Surprisal and Entropy Reduction as Measures of Cognitive Effort

Surprisal (or self-information) is a metric that quantifies the informativity of a linguistic unit, say a word, in bits of information (Hale, 2001; Levy, 2008). Intuitively, surprisal measures how predictable a word is given what we have heard or seen thus far in the sentence, with less predictable words resulting in higher surprisal values. This is captured by Equation 2.4 below, where $P(w_i|w_{0\cdots i-1})$ is the *conditional probability* that the current word w_i will follow the prefix $w_{0\cdots i-1}$.

$$Surprisal(w_i) = -log_2 P(w_i|w_{0\cdots i-1}) \tag{2.4}$$

Hale (2001) proposed that difficulty associated with processing certain syntactic structures (e.g., 'The horse raced past the barn fell', which at 'raced' is ambiguous between a main verb reading and a reduced relative clause reading) is predicted by the likelihood of the next word, as determined by an incremental syntactic parser augmented with probabilities. Hale's (2001) account was further refined and developed into Surprisal Theory (Levy, 2008), which suggests that the cognitive effort associated with processing word w_i is proportional to $Surprisal(w_i)$. In other words, the effort that comprehenders expend in processing any word in a sentence can be estimated based on the expectancy of that word in its context. Take for example the sentences in (1) below: Even though the word *stamp* is a possible continuation in both of them, it is more expected in the context of sentence 1a. Therefore, according to Surprisal Theory the cognitive effort for processing *stamp* should be higher in sentence 1b compared to sentence 1a.

(1) a. She mailed the letter without a **stamp**.
b. She went to the shop to buy a **stamp**.

Studies using reading times and electrophysiological measures have indeed established that predictability inversely affects processing effort (e.g., DeLong et al., 2005; Federmeier and Kutas, 1999; Van Berkum et al., 2005). Crucially, because surprisal is measured in the logarithmic scale, Surprisal Theory specifically predicts that small differences between low-probability words will induce large differences in cognitive effort. Such a logarithmic relationship between reading times and processing effort was shown by Smith and Levy (2013), who used linguistic corpora rather than sentence completion (cloze) tests to derive their probability estimations. Frank et al. (2015) used a similar method to investigate the influence of word surprisal on comprehenders' brain responses, and showed that surprisal indeed modulates the N400 component , previously shown to be affected by cloze probability (Dambacher et al., 2006; Federmeier and Kutas, 1999; Kutas and Federmeier, 2009).

Surprisal Theory does not make claims about the grammatical representations and mechanisms that underlie linguistic processing as observed in empirical studies. Instead, surprisal mediates between theories and observations, acting as a 'causal bottleneck' (Levy, 2008), as shown schematically in Figure 2.3. In a recent study, for instance, Delogu et al. (2017) found empirical evidence in support of a surprisal-based explanation of complement coercion, the phenomenon where event-selecting verbs, such as '*begin*' take as complements entity-denoting nouns, such as '*book*' (cf. 'John began the book' vs. 'John began writing the book'). Specifically, they found that coerced and neutral verbs (e.g., *began* and *bought*), whose (matching) surprisal values were larger than those of preferred verbs (e.g., *wrote*), elicited a larger N400 effect

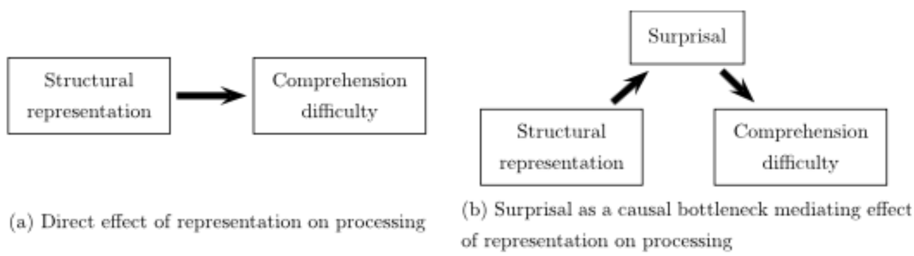


Figure 2.3 Surprisal as a causal bottleneck (Levy, 2008, p. 1133).

compared to preferred verbs. Thus, the cognitive effort associated with processing coerced nouns was shown to be due to the contextual predictability of the noun rather than some underlying type-coercion mechanism. In other words, the observed effects can be attributed to word surprisal, rather than some specialised grammatical mechanism. Thus, Surprisal Theory offers a more parsimonious account of sentence processing in this case. In sum, there is reliable evidence that cognitive effort for processing a word in a sentence can be quantified as the inverse log-probability of that word in its linguistic context.

Alternatively, the informativity of a word can be instantiated as the *reduction of uncertainty* about the possible sentence continuations that this word induces. Informativity of word w_i, measured in bits of information, can then be quantified as ΔH_{w_i}, given in Equation (2.5) below, which captures the reduction in entropy between two consecutive states in a sentence, before and after the occurrence of w_i.

$$\Delta H_{w_i} = H_{w_{i-1}} - H_{w_i} \tag{2.5}$$

Hale's (2006) Entropy Reduction Hypothesis links *entropy reduction* to processing difficulty, suggesting that the effort associated with processing word w_i should be directly proportional to ΔH_{w_i}. According to this hypothesis, comprehenders should experience some difficulty at each entropy reduction point (i.e., on every word in a sentence), but they should encounter greater difficulty the more bits of information this word reduces. This prediction was tested with reading times, both using corpora (Frank, 2010, 2013; Wu et al., 2010) and in a self-paced reading experiment (Linzen and Jaeger, 2016). In all cases, results showed that the rate of entropy reduction induced by a word was a significant predictor of processing difficulty on that word, with higher entropy reduction resulting in longer reading times.

Both surprisal and entropy reduction are calculated based on the probability distributions at a transition point, i.e., before and after the appearance of the current word in the unfolding sentence. Crucially, the transition between the 'before' and 'after' distributions is estimated in a different way for the two measures. While *entropy reduction* is calculated as the *difference* between the two distributions (see Eq. 2.5), *surprisal* is their *negative log ratio*.[2] Conceptually, this means that the two measures

[2]It can be shown mathematically that surprisal at word w_i is equal to the KL divergence between the probability distributions at w_i and $w_i - 1$ (Levy, 2008).

differ in how they instantiate the notion of word informativity: Surprisal measures how expected the current word was given the context, while entropy reduction measures what effect the current word had on reducing comprehenders' uncertainty about what is being communicated.

Clearly, surprisal and entropy reduction do not produce the same information values. As a result, they may generate different predictions regarding incremental processing difficulty (see Hale, 2016). For instance, Frank (2013) found that both measures yielded similar results when the lookahead distance used for estimating entropy reduction was kept small. The greater the lookahead distance, the weaker the correlation between surprisal and entropy reduction. These results are in line with findings from Linzen and Jaeger's (2016) self-paced reading study, where entropy reduction was calculated over the entire sentence.

Both surprisal and entropy reduction are used as estimates of processing effort, thus acting as linking hypotheses between theories of language comprehension and observations of cognitive load. They are, therefore, agnostic regarding what the mechanisms and representations underlying language processing might be (Hale, 2001, 2016; Levy, 2008). Both measures were proposed at the computational level (Marr, 1982), explaining how we derive an estimate for a word's processing difficulty from an estimate of this word's probability of occurrence. However, these theories do not make any predictions about what is taking place at the algorithmic level, and whether the two measures are related to different processing mechanisms. Two studies tackling this question (Frank, 2013; Venhuizen et al., 2019) find no evidence of distinct underlying mechanisms, and suggest that surprisal and entropy reduction rather capture different aspects of the same cognitive process.

To summarise, previous work has associated context-sensitive word informativity with processing difficulty. Two theory-neutral metrics are generally used to quantify this probabilistic notion of word informativity, surprisal and entropy reduction. Surprisal quantifies the expectancy of a word in its context, while entropy reduction the degree to which a given word decreases uncertainty about what is being communicated. Even though these measures are calculated differently and may, therefore, generate different predictions regarding incremental language processing, there is no evidence that they are derived from different comprehension mechanisms.

2.2.2 Uniform Information Density

The Surprisal Theory and the Entropy Reduction Hypothesis provide estimates of processing difficulty during comprehension. They do not, however, make any predictions regarding *production* processes. Another hypothesis, the Uniform Information Density (UID) hypothesis (Jaeger, 2010; Levy and Jaeger, 2007), links comprehension and production processes in building an information-theoretic framework for explaining speakers' choices (see also Aylett and Turk, 2004; Fenk-Oczlon, 2001; Genzel and Charniak, 2002, for similar proposals).

The UID proposes that due to cognitive resource limitations, peaks in the amount of information conveyed by words in an utterance can increase comprehenders' processing effort during word-to-word incremental interpretation. As a consequence, speakers' production choices are motivated by an intent to distribute this information (and thereby comprehenders' processing effort) as evenly as possible across their utterances. In this sense, the UID connects comprehension and production processes, attributing audience-design motivations to speakers' preferences.

More specifically, the UID predicts that speakers will prefer utterances that distribute information more uniformly across the linguistic signal. This strategy generally aims at avoiding peaks in information density (information units per time units), because they may exceed the channel capacity (see Section 2.2 above). To illustrate this, the UID predicts that, when confronted with two alternative, meaning-preserving encodings, as in Figure 2.4, speakers will prefer the one that keeps information density uniformly close to channel capacity (right panel), thereby ensuring that there will not be any points in the utterance where the listeners' comprehension system is overloaded with information (left panel).

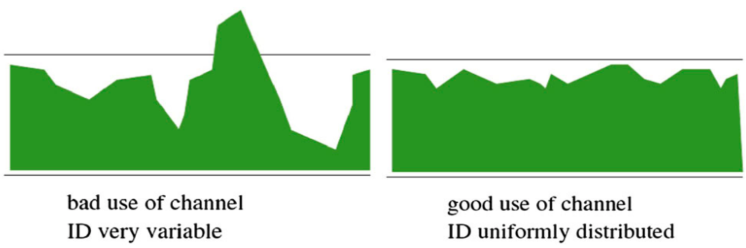


Figure 2.4 Uniform and non-uniform information transmission (Crocker et al., 2016, p. 78).

There is, indeed, evidence that speakers' choices at the (sub-)lexical level may be governed by such information distribution concerns. For instance, syllables with higher contextual predictability tend to be acoustically reduced compared to less predicted syllables, which carry more information (Aylett and Turk, 2004; Bard and Aylett, 2004). At the word level, Mahowald et al. (2013) have shown that, when both reduced and full word forms are available (e.g., *math* vs. *mathematics*), speakers are more likely to choose the reduced form when the previous context is predictive (e.g., 'She was bad at algebra, so she hated...' vs. 'She introduced herself to me as someone who loved...'). That is, reduced forms tend to be used more frequently when they are less surprising, i.e., when they convey less information. To put this idea into a different perspective, linguistic units that are more informative are not likely to be contracted, because this might result in a peak of information density in the signal.[3]

Jaeger and colleagues (Frank and Jaeger, 2008; Jaeger, 2010; Levy and Jaeger, 2007) have extended these predictions to syntactic production, and explored whether speakers' choices between alternative encodings reflect an intent to distribute information more uniformly across the entire utterance. This research has shown that choices regarding the contraction of an auxiliary verb, such as *have* (Frank and Jaeger, 2008), or the omission of the optional *that* in complement or relative clauses (Ferreira and Dell, 2000) are associated with the information – in bits – carried by the reducible words, in an effort to distribute information evenly across the utterance. More specifically, in a corpus study Jaeger (2010) showed that the production of the optional *that* complementizer is predicted by the information contained in the onset of the complement clause (CC), even when controlling for effects predicted by other accounts. In other words, it was found that the more surprising the CC-onset, the more likely the use of an overt complementizer. The informativity of the CC onset was estimated based on the subcategorisation preferences of the matrix verb: the less probable a CC was for a verb, the higher the information value of the CC onset. This is illustrated in Figure 2.5, where a CC is highly predicted after *think* in (b), but not after *confirmed* in (a). The realisation of the complementizer (see blue font characters in Fig. 2.5) allows to distribute the information contained in the CC onset across two linguistic units, thereby avoiding a peak in information density (see red font characters in Fig. 2.5).

[3] The bits of information carried by the reduced units would need to be taken over by neighbouring units, thereby increasing their information value.

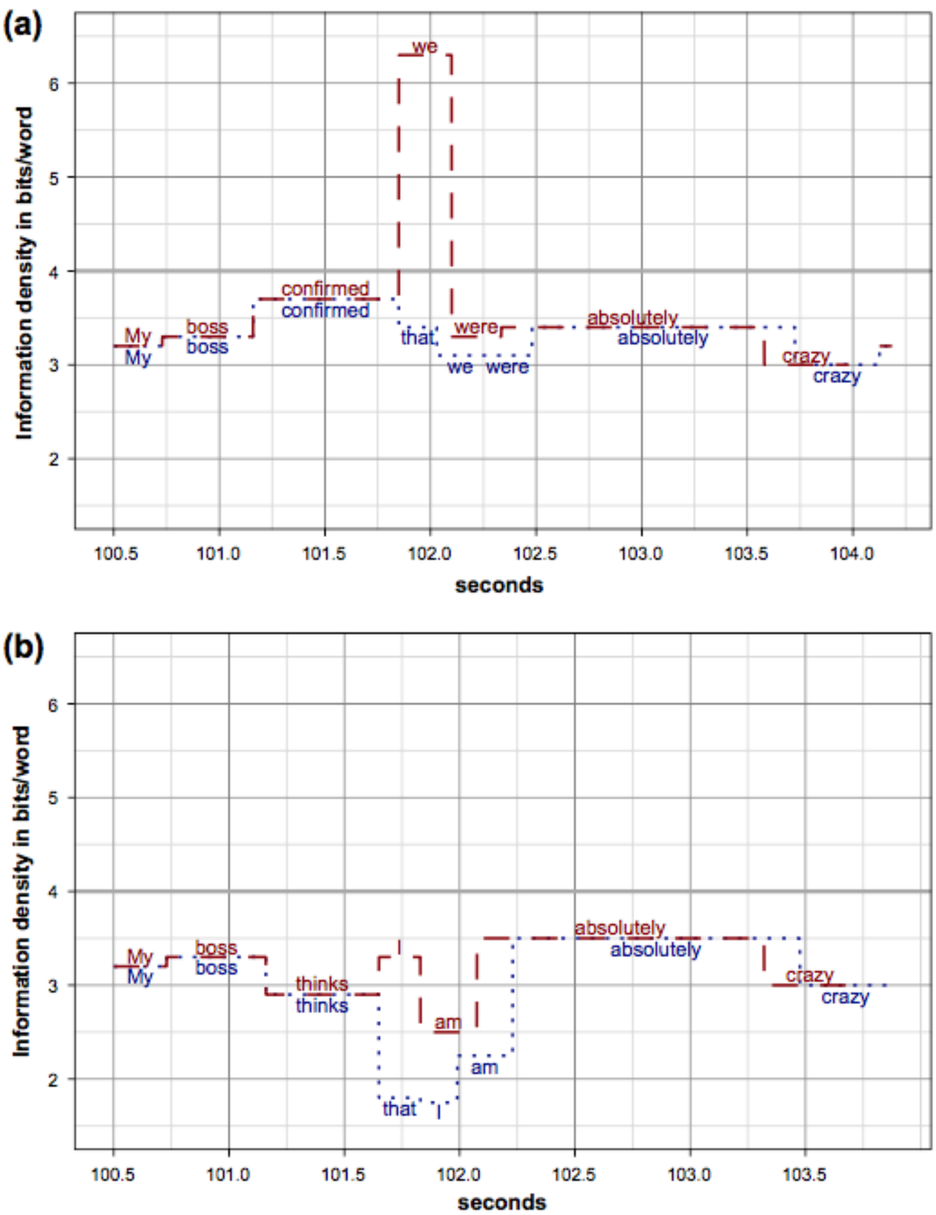


Figure 2.5 Distribution of information density across alternative utterances encoding the same message (Jaeger, 2010, p. 27).

In sum, the UID assumes that language production is informationally-sensitive, and predicts that speakers will select encodings – within the bounds of constraints defined by grammar – that distribute informationally dense content across more linguistic units. Speakers can, thus, avoid excessive processing demands in their utterances. In this sense, the UID offers a formalisation of efficient language production, linking probability-sensitive online production to probabilistic accounts of incremental language interpretation.

2.3 Extending Information-Theoretic Metrics into Situated Communication

As we saw in Chapter 1, in visually-situated communication the visual and linguistic contexts similarly influence listeners' expectations for the upcoming linguistic material in an unfolding utterance (e.g., Altmann and Kamide, 1999; Knoeferle et al., 2005; Tanenhaus et al., 1995). For example, when hearing '*Find the blue* . . . ' while immersed in a referential context such as the one in Figure 2.6, listeners may expect either of two objects to be mentioned next, the ball or the oven mitt. Therefore, the predictability of the upcoming word is, in this case, defined by the visually co-present context, as much as it is defined by the preceding linguistic context. Another way to think about this visually-determined predictability is that in the context of Figure 2.6, a listener can make predictions about what the next word is going to be based on the fact that '*blue*' in '*Find the blue* . . . ' restricts the set of potential referents to only 2 objects. That is, '*blue*' reduces listeners' uncertainty about the target referent.

We dub the uncertainty about the identity of the target referent *referential entropy*. As shown above, Shannon's entropy (Shannon, 1948) is used to quantify uncertainty about what is being communicated in a message. Entropy is calculated as shown in Equation 2.2, repeated below for convenience,

$$H(X) = -\sum_{x \in X} p(x) log_2 p(x) \tag{2.6}$$

and can be expanded to quantify referential entropy as follows: In the visual context of Figure 2.6, at 'Find the. . . ' (i.e., before any information about the target referent

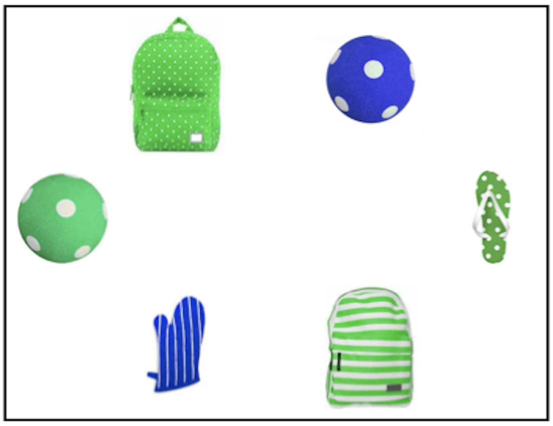


Figure 2.6 Referential entropy reduction – Example visual scene combined with the expression 'Find the blue ball'.

becomes available), all six objects are equally likely to be mentioned, and referential entropy is, therefore, equal to

$$H(X) = \sum \frac{1}{6} log_2 \frac{1}{\frac{1}{6}} = log_2 6 = 2.58 \quad (2.7)$$

bits of information. For communication to be successful, the speaker must transmit enough information to reduce referential entropy to zero. In other words, the listener's mental representation of what the target referent is must move from a state of maximum uncertainty to a state of minimum uncertainty, so that by the end of the utterance he will be able to unambiguously identify this object. As the referring expression unfolds over time, incoming words (potentially) contribute to the reduction of referential entropy. As shown above, this entropy reduction is measured by ΔH, repeated in Equation 2.8,

$$\Delta H_{w_i} = H_{w_{i-1}} - H_{w_i} \quad (2.8)$$

and it is the difference in referential entropy between two consecutive states of the listener's mental representation (or two consecutive words in the utterance, $w_i - 1$ and w_i). That is, when 'Find the blue . . . ' is uttered in the context of Figure 2.6, referential entropy at '*blue*' is 1 bit, and '*blue*' reduces entropy by $\Delta H_{blue} = 1.58$ bits. On the other hand, if the referring expression is 'Find the green . . . ', referential entropy at *green* would be 2 bits, and *green* would contribute to the reduction of entropy by ΔH_{green} = 0.58 bits. In other words, while the prenominal adjective contributes to the reduction of referential entropy in both cases, it does so to differing degrees, depending on the size of the referential domain that each adjective selects. Thus, in visually-situated communication, the information conveyed by a word does not depend only on its probability to occur in a particular (visual and linguistic) context (surprisal), but also on the amount of uncertainty about the target referent that this word reduces. By extending the Entropy Reduction hypothesis into visual contexts, cognitive effort for processing a word should be proportional to the *reduction of referential entropy* that this word induces in bits.

Under a bounded-rational approach, redundancy may be utilised as a means to distribute informationally dense content across more linguistic units. In other words, speakers may choose to encode redundant adjectives in their descriptions, when bare nouns would otherwise contain a lot of information in terms of reducing referential entropy. For instance, the minimally-specified '*the mitt*' or the over-specified '*the blue mitt*' in the context of Figure 2.6 both successfully establish reference, reducing the initial entropy (H_{find_the} = 2.58 bits) to zero. However, this reduction takes place in a single word in the minimally-specified expression, while it is distributed across two words when the expression is over-specified (cf. information density at *we* vs. *that* in Fig. 2.5a). That is, redundancy distributes information across a longer sequence of words, thus strengthening the signal while also providing listeners with additional cues to guide their visual search, making referent identification faster, less effortful, and more reliable.

In sum, we propose an extension of information-theoretic measures, such as entropy reduction, into visually-situated communication, in order to quantify the effort associated with referential processing, which may in turn motivate production choices.

Chapter 3

The online processing of over-specifications

Chapter 2 highlighted the conflicting evidence regarding the comprehension of over-specifications: While some studies suggest that over-specifications hinder comprehension (cf. Davies and Katsos, 2013; Engelhardt et al., 2011, 2006), others find that they have a facilitatory effect (cf. Arts et al., 2011a; Brodbeck et al., 2015). This evidence, however, comes from experiments that vary in important factors, such as the size of the referent set, or the kind of adjectives used, possibly contributing to the observed discrepancy among the obtained results. Additionally, earlier work has often employed offline dependent measures, such as acceptance ratings, which makes it impossible to draw conclusions regarding the online processing of over-specifications. The goal of Experiment 1 was to (a) determine whether over-specification hinders or facilitates comprehension, and (b) identify the neural underpinnings of referential processing in visual contexts.[1]

3.1 Experiment 1

In Experiment 1, we assessed whether the presence of a redundant prenominal adjective incurs a processing cost, as the Gricean account would predict (see Section 2.1). According to this account, adjectives should only be used when necessary to

[1]The study reported in this chapter was published in the conference proceedings of CogSci 2015 (Tourtouri et al., 2015).

distinguish a target referent from objects of the same type, by encoding a property in which they differ. Therefore, listeners should expect adjectives to be used only in reference to objects that belong in contrast pairs, and should interpret adjectives contrastively.

Furthermore, while it is uncontroversial that referential ambiguity disrupts comprehension, as listeners judge under-specified expressions to be worse than minimal descriptions (e.g., Davies and Katsos, 2013; Engelhardt et al., 2006), the online processing of under-specifications has been scarcely examined. Some ERP studies have established a neural marker of referential ambiguity in discourse, the Nref effect (Nieuwland and Van Berkum, 2008a,b). The Nref is a frontally-distributed sustained negativity associated with nouns that are ambiguous between potential antecedents in prior discourse. In situated contexts, (indirect) evidence from an ERP study on online perspective taking (Sikos et al., 2019a) suggests that referential under-specification also manifests as an Nref effect for the (under-specified) noun. To our knowledge, however, no previous research has *directly* assessed listeners' neural responses to under-specification in visually-situated comprehension. Therefore, the secondary goal of this study was to explore the neurophysiological index of under-specification in situated comprehension.

Participants' EEG was recorded as they attended to spoken instructions like 'Find the yellow bowl' in German coupled with visual displays such as the ones in Figure 3.1. The combination of one instruction with four visual scenes generated four experimental conditions. In the minimally-specified condition (MS), which served as a baseline, the adjective was necessary to identify the target referent, because a second object of the same type but different colour was also present (see Fig. 3.1a). In the over-specified condition (OS), 'yellow' was unnecessary, because the bowl was singleton (see Fig. 3.1b) and could, therefore, be identified by mention of its type alone. In the under-specified condition (US), the adjective was not sufficient for target identification, as it selected two objects of the same type and did not disambiguate between them (see the yellow bowls in Fig. 3.1c). Finally, we also included a mismatch condition (MM), where the object identified by the adjective (the yellow mug in Figure 3.1d) did not match the one mentioned by the noun (the bowl in Fig. 3.1d). The MM condition served as a negative baseline – a case of *explicit* referential failure – balancing the experimental design.

The experiment employed a referential processing task (Krauss and Glucksberg, 1977; Krauss and Weinheimer, 1964), combining spoken sentences with visual scenes.

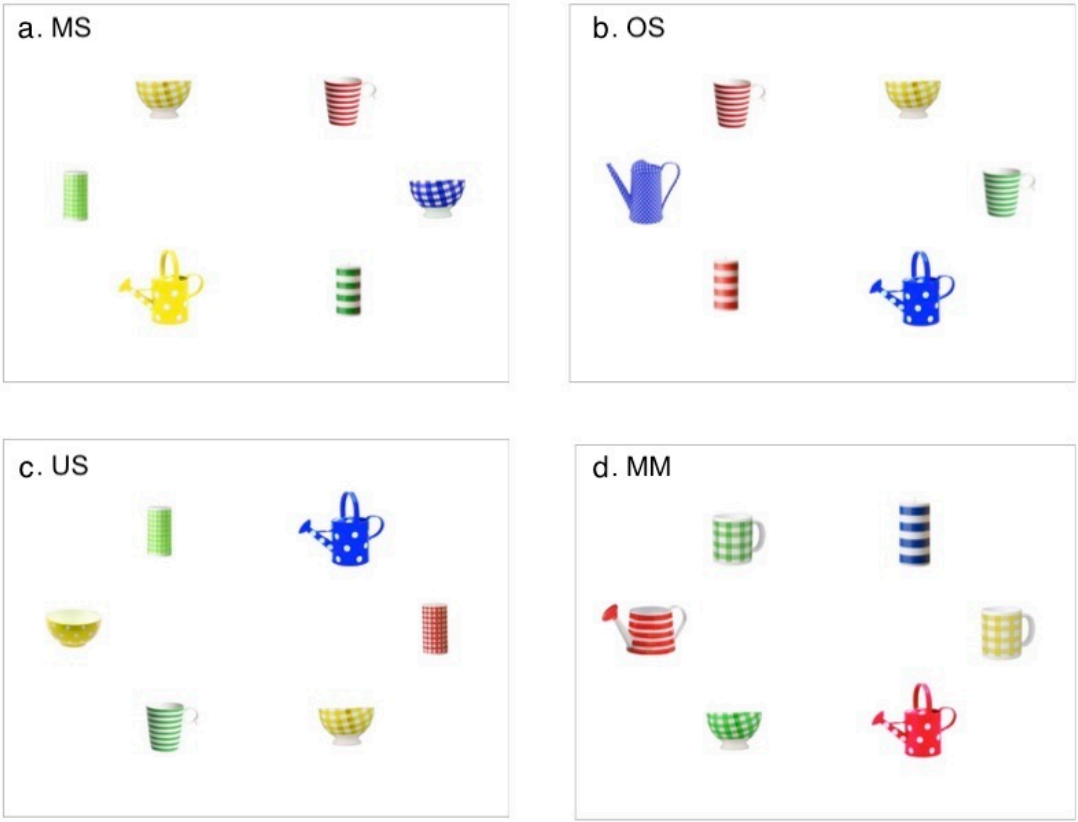


Figure 3.1 Experiment 1. Sample visual scenes for a colour experimental item, paired with the audio instruction 'Find the yellow bowl'.

In Engelhardt et al. (2011), visual scenes always presented two objects differing in a maximum of two features (recall Fig. 2.1). Consequently, their results may be limited to such simple contexts, in which redundancy is striking and may lead to attempts at pragmatic inferencing. In this experiment, we therefore increased the number of referential candidates as well as the number of dimensions across which objects could differ in any one visual context (colour, pattern and type).

According to the Gricean account, processing should be more difficult in the OS condition compared to the MS condition. This difficulty may be indexed by an N400 effect for the OS relative to the MS condition, which may arise already after adjective onset (see Engelhardt et al., 2011), since at this point listeners can already detect that the adjective is not required to select the target referent. If, on the other hand, Gricean concerns are not at play in referential processing, no such difference should be observed.

The US condition should result in a processing cost relative to the MS condition, because in the US condition the expression fails to establish reference. What is interesting, however, is what kind of effect this comparison may give rise to. One possibility is that this cost may manifest as an Nref effect (Nieuwland and Van Berkum, 2008b; Sikos et al., 2019a) reflecting referential ambiguity. This effect may emerge on the adjective, as at this point it is already apparent that the adjective does not help to discern between the two same-type referents (cf. the two bowls in Fig. 3.1c; a pattern adjective would minimally distinguish between them). Alternatively, we may observe an increased N400 on the noun for the US vs. the MS condition, as after the adjective listeners might still expect a second adjective to occur in the US condition, disambiguating between the two potential referents. (Filler items were included that used two pre-nominal adjectives to distinguish between these objects).

The MM condition served as a negative baseline. We were primarily interested in the comparison between the MM and the US condition, in order to establish whether referential failure due to under-specification is similar to explicit referential failure due to the mismatch between the linguistic and visual input. In the MM condition, a specific referent is expected after hearing the adjective (cf. the mug in Fig. 3.1d), but the noun selects another one (cf. the bowl in Fig. 3.1d). This unexpected noun should yield a larger N400 in the MM condition relative to the MS condition. What is of greater interest, however, is whether the US condition will result in a response similar to that in the MM condition, as they both fail to select a referent, or a qualitatively different one, reflecting the different nature of this failure.

3.1.1 Method

Participants. Thirty-three students (average age = 25, 25 female) from Saarland University participated in the experiment after giving written informed consent, and were monetarily compensated for their participation. All participants were right-handed, native speakers of German with normal or corrected-to-normal vision, and no problems with colour perception.

Materials.

Pictures of everyday objects (e.g., mugs, bowls, etc.) were used to create the visual stimuli. The objects differed in colour (red, blue, yellow, green) and pattern (chequered, dotted, striped). Pattern was used instead of size, which is more commonly used in such tasks (cf. among others Engelhardt et al., 2011; Sedivy et al., 1999), because pattern – like colour but unlike size – is an intrinsic property of the object and does not invoke a comparison to other potential referents in the visual context. In this way we ensured that any preference for a Gricean (i.e., contrastive) interpretation of the adjective would be due to our manipulation and not due to the contrastive nature of size adjectives. GIMP (Version 2.8.10) was used to adjust the colour hue and brightness of the object pictures. The pictures were then submitted to an offline picture naming study measuring naming agreement for the objects. Twenty-four independent participants were presented with the object pictures in all colours and patterns (distributed across eight lists), and were asked to provide descriptions including colour and pattern. Thirty-two objects with inter-rater agreement 80% or higher were then employed to create the visual stimuli (see Appendix A for a full list of the objects in the different colours and patterns).

In total, 640 visual stimuli were created, of which 512 were used to create the experimental items, 128 were used for the fillers and 12 for practice trials. Experimental items were the combination of four displays and a single spoken instruction (cf. Fig. 3.1). This gave rise to 128 items, half of which were paired with colour instructions (colour items) and the other half with pattern instructions (pattern items).

All displays were created in a way that neither the target feature nor the target referent were identifiable before hearing the adjective and noun in the accompanying instruction. To this end, six objects were used per display: two pairs of objects for the MS and US conditions, and two singletons for the OS and MM conditions. Figure 3.1a shows the display for the MS condition, where the target referent (the yellow bowl) belongs in a contrast pair with an object of the same type and pattern but different colour (the blue bowl), thereby making the use of a colour adjective necessary for its identification. Four distractor objects fill the remaining positions in the colours and patterns that would allow these objects to function as targets in the other three conditions. That is, the red mug could potentially be the target referent in an OS condition, as 'mug' would suffice for target identification and the use of an adjective would be redundant. The two green candles could potentially be referents in an US condition, since 'green' is neither necessary nor sufficient for disambiguating between the two candles. Lastly, the yellow watering can could be the target in a MM condition. The rest of the colour item displays were designed in the same way, with the display structure allowing all objects to be potential targets in different conditions.

Pattern item displays were created in a similar way, with pattern being the mentioned feature (see Fig. B.1 in Appendix B for an example pattern item). That is, in the MS condition the target referent differed from its competitor in the mentioned feature; in the OS condition the target was the single object in the display bearing the mentioned feature; in the US condition there were two objects bearing the mentioned pattern but differing in colour; in the MM condition, the referent did not carry the pattern mentioned in the expression, even though there was one object available with this feature.

Because determiners in German are marked for gender, only same-gender objects were used in experimental displays, thus ensuring that the determiner would not reveal the target referent. Similarly, no phonological competitors (e.g., *Schüssel - Schürze*) appeared in the same scene, so that the adjective onset would always be the earliest point of disambiguation.

The apparent inconsistency between colour and pattern items (i.e., that 4 colours were present in colour displays, while only 3 colours were present in pattern displays) was counterbalanced in the fillers. Specifically, in filler items 3-colour displays were paired with colour instructions and 4-colour displays with pattern instructions. Hence, across trials, before adjective onset, there were no cues leading up to the target object, the condition or the distinguishing feature. Fillers also counterbalanced the target position, as in experimental trials the target referent always occupied one of the four innermost positions. Therefore, in filler trials target objects occupied each position as many times as was necessary to make sure that across trials every object appeared as target in each position equally frequently. What is more, instructions in filler trials were the MS counterparts of the OS and US experimental instructions. That is, in half of the fillers we used the MS versions of the OS critical instructions (e.g., 'the bowl' instead of 'the yellow bowl' in Fig. 3.1b); the other half of the fillers were based on the US critical trials, but always contained adequate information for identifying the target referent (e.g., 'the yellow chequered bowl' instead of 'the yellow bowl' in Fig. 3.1c). Finally, we counterbalanced the target colour and pattern, making sure that across all trials every object appeared as target referent in all four colours and three patterns equally frequently.

The audio stimuli were recorded with neutral intonation by a female native speaker of German in a sound isolated cabin using Cubase AI5. All instructions started with the words '*Finde den/die/das...*' ('Find the. . . ' with the definite article in accusative in the masculine, feminine and neuter gender). Critical instructions continued with one prenominal adjective mentioning either colour (*gelbe*) or pattern (*karierte*), followed by the head noun (*Schüssel*), while filler instructions had two or zero adjectives. As speech was continuous, recordings were later annotated for adjective and noun onsets and durations using Praat (Version 5.3). Mean adjective duration was 481.3 ms (*SD* = 32 ms) and mean noun duration 557.2 ms (*SD* = 75.7 ms).

Stimuli were divided into 4 lists of 256 trials using a Latin Square design, so that only one version of an item was on each list and no one participant saw more than one condition of a given item. Lists were pseudo-randomized for each participant, so that no more than two experimental items were consecutive, and that, even when a filler interfered, trials of the same condition would not be adjacent. The experiment was implemented and run using the E-prime software (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Procedure. Participants were seated in a sound-isolated and electromagnetically shielded cabin at a comfortable distance from a 1680 x 1050 resolution monitor. After they read the instructions for the experiment, they were presented with displays of each object in all colours and patterns accompanied by pre-recorded audios of each objects' type. Before starting the experimental session, participants were presented with 12 practice trials, which aimed at familiarising them with the experimental task. The task was to indicate whether the target referent appeared on the left side or on the right side of the screen (MS and OS conditions), or whether it was not possible to determine its position (US and MM conditions). Specifically, participants were asked to press one of four buttons to indicate their response: two buttons labelled LEFT and RIGHT indicating the side of the screen where the target referent appeared on, and two buttons with a question mark indicating that it was not possible to specify the target objects' location (see the prompt screen in Fig. C.1 Appendix C). A correct response in the MS and OS conditions would require participants to press either the LEFT or the RIGHT button, while a correct response in the US and MM conditions would require them to press one of the two question mark buttons (either one of them). When it was made sure that the participant understood the task, the experimental session began. One session was divided into 8 blocks of 32 trials, in between which participants could take short breaks, and its average duration was 70 minutes.

A trial started with a 2.5 seconds preview of the visual scene, as is shown in Figure 3.2. After this time, a cross appeared in the middle of the screen, and 500 ms later the spoken instruction started. At the offset of the audio stimulus the cross disappeared, while the objects remained on the screen for a wrap-up period of 500 ms. Participants were required to fixate the cross while it remained on the screen. In each trial, the display was visible for a total of approximately 5 seconds. At the end of each trial, participants saw a prompt screen reminding them to perform the task. Responses were given in the form of button presses in a Cedrus response pad (Cedrus Corporation, San Pedro, California, USA).

Participants' EEG was recorded from 26 Ag/AgCl electrodes placed on the scalp according to the standard 10-20 system. The EEG signal was amplified by a BrainAmps DC amplifier (Brain Products, GmbH, Munich, Germany) and digitised at a sampling rate of 500 Hz. Eye movements and blinks were monitored by

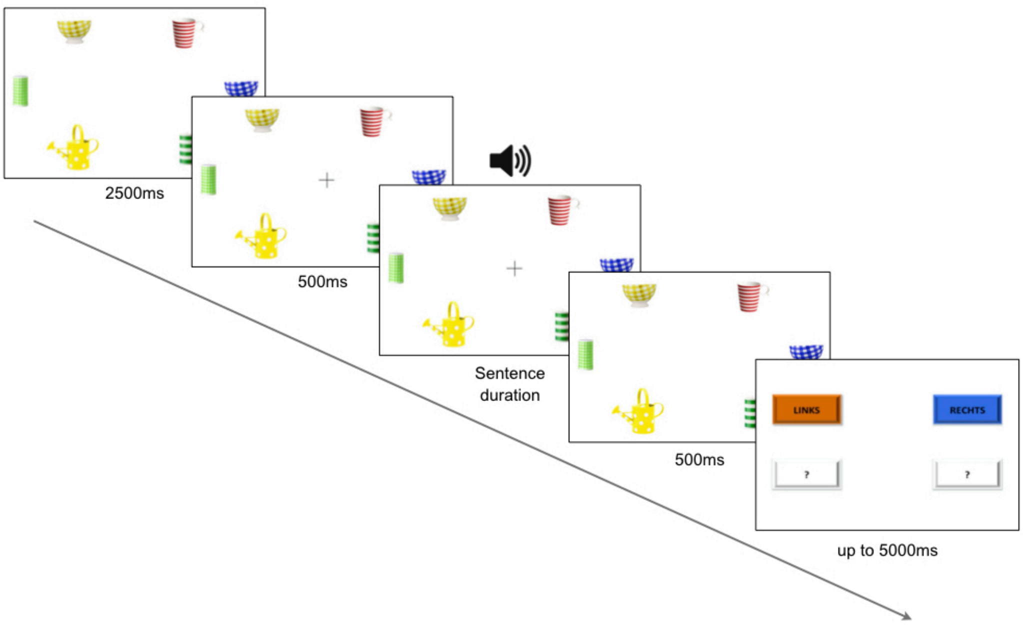


Figure 3.2 Experiment 1. Example trial sequence and timing of events.

electrodes placed on the outer canthus of each eye, and above and below the right eye. Impedances were kept below 5 kΩ.

Analysis. The offline processing of the EEG data was performed using BrainVision Analyzer 2 (Brain Products, GmbH, Munich, Germany). The EEG signal was filtered (30 Hz high cut-off) and re-referenced offline to the average of the two mastoid electrodes. Single-participant averages were then computed in a 1000 ms window per condition relative to the onset of the adjective (*gelbe*) and head noun (*Schüssel*), and aligned to a 200 ms 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 per condition were excluded from the analyses. Only artefact-free ERP averages time-locked to the onset of the critical regions entered the analyses.

We report statistics for response times (RTs) and ERPs. For the analysis of RTs, we fitted linear mixed models (lme4 package; Bates et al., 2015) in R (version 3.5.1; R Core Team 2018). *Specificity* of the referring expression was treatment-coded with the MS condition as the baseline level, and was included as a fixed factor in the model. Crossed random intercepts and slopes for participants and items were also included in the model. For the ERPs, we proceeded to planned pairwise comparisons between conditions (OS vs. MS, US vs. MS, MM vs. MS). We carried out analyses in the adjective region, where Engelhardt et al. (2011) report results. Since a redundant prenominal adjective may affect comprehension of the subsequent noun (cf. Sedivy et al., 1999), analyses were also performed on the noun. Note that the US vs. MM comparison that we set out to investigate is relevant only in the noun region, where the noun is heard but fails to select a referent in the MM condition (which, up to the adjective, is identical to the MS and OS conditions). If the maximal model did not converge, we simplified the random effects structure as suggested by Barr et al. (2013).

For analysis of the ERPs, following Kuperberg et al. (2010) we performed ANOVAs (ez package; Lawrence, 2011) at a Midline column containing electrodes Fz, Cz and Pz, and two lateral columns: the Medial column, containing electrodes FC1, C3, CP1, FC2, C4 and CP2, and the Lateral column, containing electrodes F3, FC5, CP5, P3, F4, FC6, CP6 and P4. For the Medial and Lateral columns, within-subjects factors were Specificity (levels: MS, OS, US, MM), Longitude (levels: anterior, central, posterior), and Laterality (left, right). For the Midline column, only the factors of Specificity and Longitude were included in the ANOVAs. Where appropriate, we will report the Greenhouse-Geisser corrected p-value (Greenhouse and Geisser, 1959) with the original degrees of freedom.

3.1.2 Results

Response accuracy was very high overall. Specifically, participants gave the correct response at a rate of 90% in the MS condition, 96.1% in the OS condition, 86.8% in the US condition, and 90.4% in the MM condition. We analysed response accuracy data using generalised linear mixed models (lme4 package (Bates et al., 2015) in R (version 3.5.1; R Core Team, 2018) with a binomial function. Because the maximal model including random intercepts and slopes for both participants and items did not converge, we simplified the random effects structures as suggested by Barr et al. (2013). The Specificity factor was dummy-coded with the MS condition as the reference level. Response accuracy was higher in the OS compared to the MS condition (β = 1.153, *SE* = 0.252, z = 4.563, $p < .001$), and higher in the MS compared to the US condition (β = -0.451, *SE* = 0.202, z = -2.231, p = .026), but no significant differences were found between the MS and MM conditions (p > .05). All further analyses included trials with correct responses only.

Response times. RTs were time-locked to the onset of the prompt display, and analyses were carried out on log-transformed RTs. The results indicated that participants responded faster (β = -0.059, *SE* = 0.026, t = -2.23, p = .026) in the OS condition (439.02 ms, *SD* = 250.4 ms) and slower (β = 0.085, *SE* = 0.036, t = 2.305, p = .031) in the MM condition (540 ms, *SD* = 348.8 ms) compared to the MS condition (480 ms, *SD* = 306.4 ms). The comparison between the US (482 ms, *SD* = 356.9 ms) and the MS conditions did not result in a significant difference ($p > .05$).

ERPs. Visual inspection of the ERP waveforms time-locked to the onset of the adjective (see Fig. 3.3) indicates a larger positivity for the US condition compared to the other three conditions, starting at approximately 400 ms and extending through the noun region. The MS, MM, and OS conditions patterned together throughout the adjective region, and started to diverge only after approximately 600 ms (already into the noun).

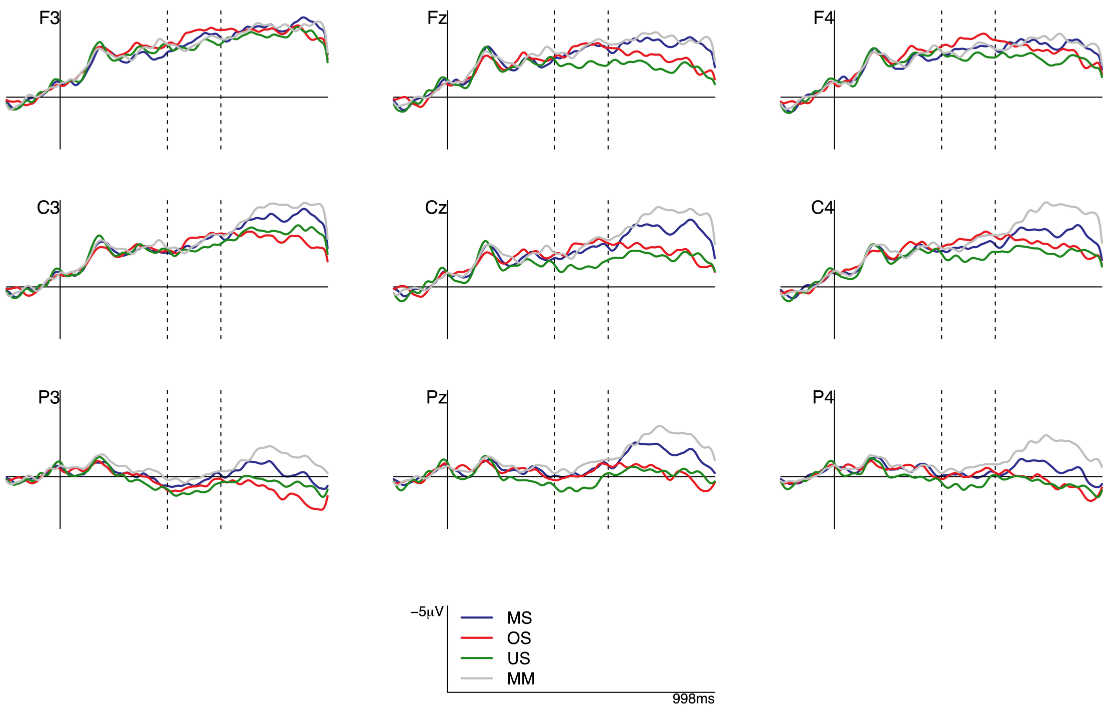


Figure 3.3 Experiment 1. Averaged ERPs time-locked to adjective onset. Dotted lines indicate the 400-600 ms analysis window. Negative voltages are plotted upward.

Because mean adjective duration was 481 ms, and therefore any later effects should be associated with processing of the subsequent noun, we performed analyses in a time-window between 400 and 600 ms (see dotted lines in Fig. 3.3). For the comparison between the US and MS conditions, the ANOVAs revealed a main effect of Specificity ($F(1,22) = 5.159$, $p = .033$, $\eta_p^2 = .029$) in the midline sites, and a Specificity x Laterality interaction ($F(1,22) = 4.2$, $p = .05$, $\eta_p^2 = .001$) in the medial sites. These results indicate that the effect was broadly distributed and slightly more pronounced in the right electrode sites (see Fig. 3.4). The comparisons between the OS vs. MS and the MM vs. MS conditions did not reach significance ($p > .05$) in this time-window.

In the noun region, we observed a graded centro-parietal negativity peaking at around 400 ms after noun onset, with the MM condition being the most negative and the OS condition the least negative (see Fig. 3.5), while the US condition elicited a positivity starting at around 400 ms. For the comparison between the OS and MS

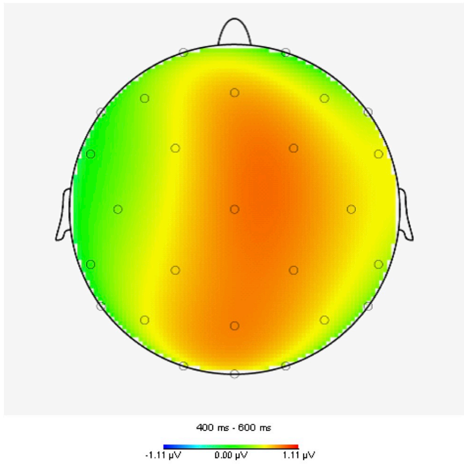


Figure 3.4 Experiment 1. Topographic map showing the difference between the US and MS conditions in the time-window 400-600 ms post adjective onset.

conditions in the 300-500 ms time-window (see Fig. 3.6a), the ANOVAs revealed a reduced N400 for the OS condition, which was significant in all midline (F(1,22) = 5.126, p = .034, η_p^2 = .056), medial (F(1,22) = 6.238, p = .02, η_p^2 = .065), and lateral (F(1,22) = 5.63, p = .027, η_p^2 = .044) electrode sites. The comparison between the MM and MS conditions in the same time-window (see Fig. 3.6b) resulted in the following effects: an interaction of Specificity x Longitude (F(2,44) = 4.392, p = .032, η_p^2 = .006) in the midline sites; an interaction of Specificity x Longitude (F(2,44) = 7.059, p = .003, η_p^2 = .004) in the medial sites; and in the lateral sites a marginal effect of Specificity (F(1,22) = 3.818, p = 0.063, η_p^2 = .043), as well as a Specificity x Longitude (F(1,22) = 11.740, p = 0.002, η_p^2 = .008) and a Specificity x Laterality (F(1,22) = 5.819, p = .003, η_p^2 = .004) interaction. These results indicate that the observed graded N400 effect was broadly distributed and slightly more pronounced over the right electrode sites (cf. Fig. 3.6). This effect suggests that referential redundancy facilitated comprehension in the presence of a complex visual context. Crucially, the OS and MM conditions were identical up to the noun: a single object in the scene matched the property mentioned by the adjective, in both cases.

While the comparison between the US and MS conditions did not reach significance in the 300-500 ms time-window, the waveforms of the US and MS conditions

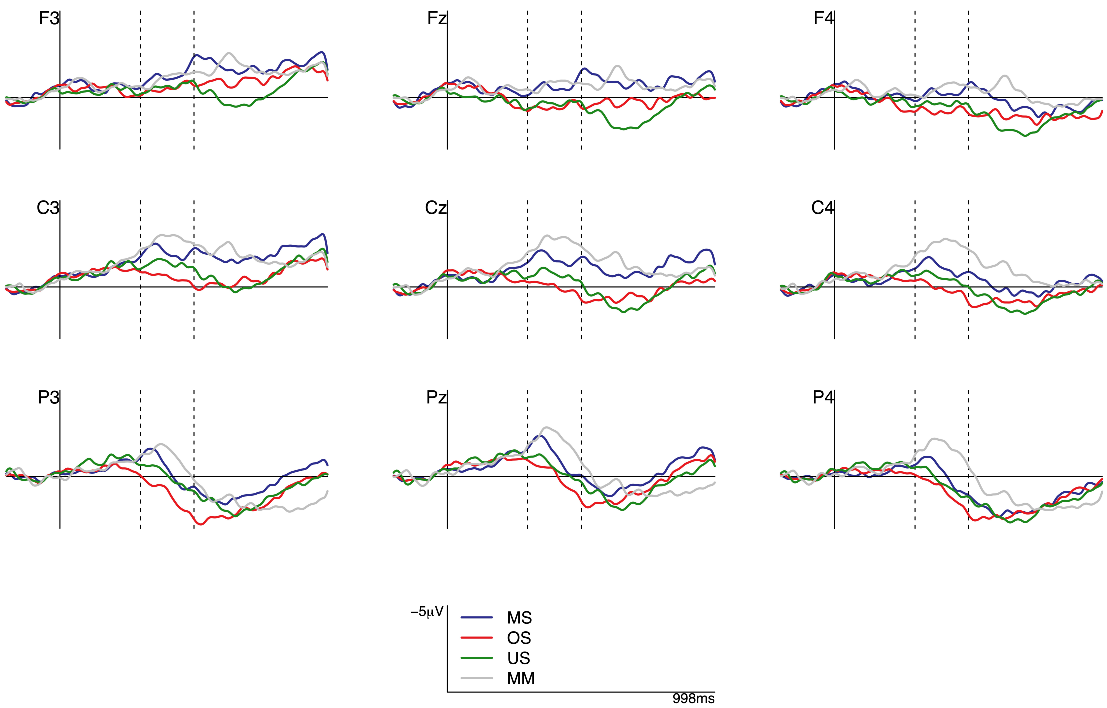
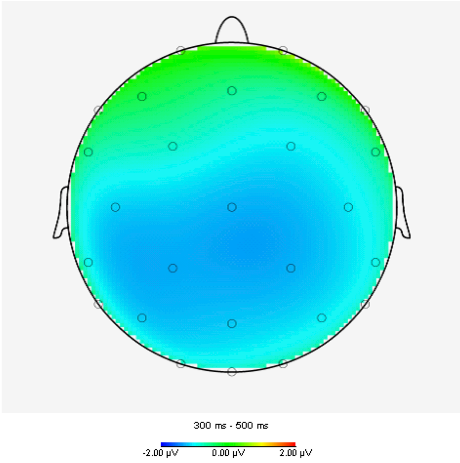


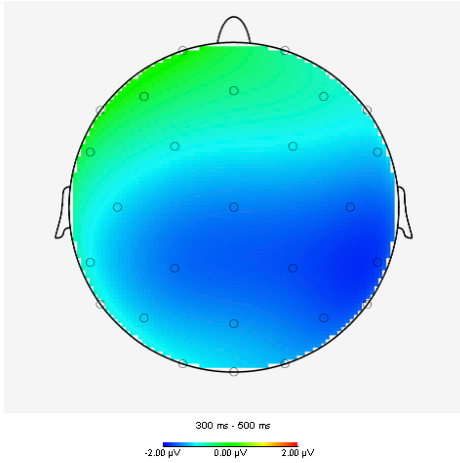
Figure 3.5 Experiment 1. Averaged ERPs time-locked to noun onset. Dotted lines indicate the 300-500 ms analysis window. Negative voltages are plotted upward.

seem to diverge at roughly 500 ms post-noun onset (cf. Fig. 3.5).[2] We, therefore, conducted further analyses in the 500-800 ms time-window. In this time-window, the comparison between the US and MS conditions revealed an interaction of Specificity x Longitude in the medial ($F(2,44) = 3.271$, $p = .047$, $\eta_p^2 = .002$), and lateral ($F(1,22) = 6.425$, $p = .019$, $\eta_p^2 = .007$) sites, while in the lateral sites the Specificity x Longitude x Laterality interaction ($F(1,22) = 7.215$, $p = .013$, $\eta_p^2 = .001$) was also significant. These results indicate a frontal positivity for the US compared to MS condition that was more pronounced over the left electrode sites. The results for the comparisons between the MM vs. MS conditions and OS vs. MS conditions were similar to those in the previous time-window (300-500 ms). Specifically, for the comparison between the OS and MS conditions we found a main effect of Specificity in the

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



(a) MS minus OS



(b) MM minus MS

Figure 3.6 Experiment 1. Topographic maps in the time-window 300-500 ms post noun onset.

midline channels ($F(1,22) = 5.126$, $p = .033$, $\eta_p^2 = .056$), and a Specificity x Longitude x Laterality interaction in the medial channels ($F(2,44) = 5.184$, $p = .009$, $\eta_p^2 = 0.001$). For the comparison between the MM and MS conditions, the ANOVAs revealed a Specificity x Longitude interaction in the midline ($F(2,44) = 4.392$, $p = .032$, $\eta_p^2 = .006$) sites, a Specificity x Laterality interaction in the medial ($F(1,22) = 5.597$, $p = .027$, $\eta_p^2 = .001$) and lateral ($F(1,22) = 12.959$, $p = .001$, $\eta_p^2 = .007$) sites, as well as a Specificity x Longitude x Laterality interaction in the medial sites ($F(2,44) = 4.312$, $p = .022$, $\eta_p^2 = .001$). These results indicate that the graded N400 effect elicited on the noun was sustained, and in the later time-window it was more prominent over right centro-parietal electrodes.

3.1.3 Discussion

Experiment 1 investigated the neurophysiological correlates of over-specified and under-specified referring expressions in visually-situated comprehension. While both kinds of expressions violate the Gricean maxims of Quantity, as neither of them provides the precise amount of information that is necessary to identify a target referent, only under-specifications have been shown by previous research to be unquestionably disfavoured by comprehenders (cf. Davies and Katsos, 2013; Engelhardt et al., 2006). However, this finding has so far mostly been shown with offline acceptability rating studies, and no previous work has directly tackled the online processing of under-specifications in visual contexts (but see Sikos et al., 2019a). As for over-specifications, previous research has found mixed evidence regarding whether referential redundancy is detrimental to processing or not (cf. Arts et al., 2011a; Engelhardt et al., 2011).

The present study sheds light into this long-standing debate by examining the neural responses to over-specified and under-specified referring expressions within a single experimental design. Specifically, in an ERP experiment we measured participants' brain responses to utterances like 'Find the yellow bowl' (in German) that were over-specified (OS) in the presence of a visual context containing one matching object, and under-specified (US) when two matching objects were available. We compared processing of over-specified and under-specified utterances to that of their minimally-specified counterparts (MS; when a second bowl of different colour was available in the scene), as well as to cases of mismatch between the linguistic and visual input (MM; when the only bowl present in the scene was not yellow).

Our results present two important insights. Firstly, they indicate that referential over-specification is beneficial rather than detrimental to comprehension – at least when the visual scene is demanding – as indexed by the decreased N400 effect elicited for the OS relative to the MS condition. This interpretation is corroborated by behavioural measures, which show that participants were faster to locate the target in the OS compared to the MS condition. Secondly, we found that, in contrast to under-specification in discourse contexts, which yields an Nref effect (Nieuwland and Van Berkum, 2008a,b), under-specification in visual contexts results in a long-lasting positivity starting at around 400 ms after the onset of the ambiguity (adjective) and extending through the end of the subsequent region (noun), where the ambiguity is still unresolved. This effect was qualitatively different from what observed for cases of explicit referential failure due to a mismatch between the linguistic and visual input (MM condition).

More specifically, in the adjective region the ERP waveforms for the OS, MM and MS conditions overlapped, while the US condition elicited a long-lasting positivity (see Fig. 3.3). This effect was present for both US vs. MS and US vs. MM comparisons. Regarding the first one, in both conditions the adjective selected two objects, but only in the MS condition were these objects of a *different type* (cf. the bowl and the watering can in Fig. 3.1a). More information was, therefore, required after the adjective in order to identify the type of the target object and fully establish reference. By contrast, in the US condition the adjective selected two objects of the *same type* (cf. the two yellow bowls in Fig. 3.1c), and while this allowed listeners to predict the *noun* (the two bowls were the only yellow entities in Fig. 3.1c), it did not help to disambiguate *reference* – a pattern adjective would have been more appropriate.

Based on previous work on the online effects of under-specification, we should expect the comparison between the US and MS conditions to elicit an Nref effect (a frontally-distributed sustained negativity). The Nref has been largely established as the neural signature of referential ambiguity in discourse comprehension (Nieuwland and Van Berkum, 2008b), while there is recent evidence that it also indexes ambiguity relative to referents in visual contexts (Sikos et al., 2019a). However, other literature has also found under-specifications to yield positive ERP effects. Nieuwland and Van Berkum (Nieuwland and Van Berkum, 2008a), for example, report a positive deflection (late positive component; LPC) associated with referentially ambiguous anaphors for a subset of participants, in which the ambiguity did not elicit an Nref effect. The LPC was interpreted in the context of other positive components, such as the P300 and the P600, and was linked to task strategies that participants potentially adopted. Another ERP study examining the processing of partial answers (e.g., 'The mayor praised the councilor') to questions (e.g., 'What did the mayor and the alderman do?') (Hoeks et al., 2013) also tackled the issue of under-specification in discourse. Relative to a neutral condition where the question was generic (e.g., 'What happened?') and the same answer was therefore complete, partial answers resulted in a broadly distributed positivity, in a time-window 300-900 ms after the onset of the critical word (e.g., 'councillor'). This positivity was interpreted as reflecting increased effort in updating the mental representation of what is being communicated. Analogously, the positivity elicited by the US compared to the MS condition at the adjective may reflect two different processes. First, it might reflect participants' realisation that the information on the adjective was not helpful and their ensuing readiness to give the appropriate response, in case no further information became available (i.e., to push the question-mark button). Secondly, this positivity may index a process of updating the mental model of what is being communicated. This update can amount to a general expectancy for a disambiguating adjective to occur before the noun, or even the formulation of specific predictions of what that adjective should be given what is already known (cf. 'dotted' or 'chequered' in Fig. 3.1c). We return to the interpretation of this positivity in Section 4.1.3.

The US vs. MM comparison also exhibited a positive deflection in the noun region (see, however, footnote in Section 3.1.2). This is a novel finding, indicating that referential failure in the two conditions is qualitatively different. Even though the adjective in the MM condition was also unnecessary (cf. Fig. 3.1d, where 'yellow' selects a singleton referent), it did however help narrow down the set of potential referents to only one object, thus allowing the prediction of the upcoming noun and the identification of the target referent.[3] In the US condition, on the other hand, the adjective identified exactly two objects that were of the same type, thus letting listeners predict the head noun (e.g., 'bowl') but not allowing them to identify *which* bowl was the target. As discussed above regarding the US vs. MS comparison, this positivity for the US condition may reflect either a task-related effect or some kind of updating of the listeners' mental representation.

In the noun region, we found a graded centro-parietal negativity for the MM, MS and OS conditions, which peaked at around 400 ms after the onset of the noun with the MM condition being the most negative and the OS condition the least negative (see Fig. 3.5). This increased N400 amplitude for the MM condition relative to the MS condition indexes that word retrieval was hindered for the noun in the MM condition (cf. Brouwer et al., 2012), and thus reference failed to be established. On the contrary, the attenuated N400 observed for the OS compared to the MS condition reflects that the redundant adjective facilitated lexical retrieval on the noun, thereby benefiting referential processing. This finding suggests that Gricean considerations are not at play in referential processing, at least when the visual context is as complex as those used here. In contrast, the use of a redundant prenominal adjective facilitated processing of the following word.

Nonetheless, it seems that the facilitation observed for the OS condition cannot be attributed only to the use of a redundant adjective, i.e., it might not be an effect of over-specification per se. It seems that the structure of the visual scenes contributed to this effect. Crucially, contrary to the MS condition, in the OS condition the adjective (e.g., 'yellow') selected exactly one object (cf. the bowl in Fig. 3.1b), which allowed listeners to formulate *precise predictions* about the upcoming noun. This was also the case in the MM condition, as the adjective there also selected only one object (cf. the mug in Fig. 3.1d). By contrast, in the MS condition the adjective selected two objects (cf. the bowl and the watering can in Fig. 3.1a) and both of

[3]At this point, before it was falsified by the noun, this prediction was perceived as helpful, as indexed by the overlapping waveforms for the MS, OS and MM conditions on the adjective.

them could be mentioned by the noun. When the noun was finally heard, listeners' predictions were confirmed in the OS condition but cancelled in the MM condition. Therefore, the visually-determined *expectancy* of the noun was highest in the OS condition and lowest in the MM condition, and this is what the graded N400 effect reflects.[4] As lower expectancy – and thus, higher surprisal – has been associated with increased processing difficulty (see Section 2.2.1 above), this effect indicates that referential processing was facilitated in the OS condition, but hindered in the MM condition.

One question that arises in the face of these results is whether over-specification would still be found to facilitate comprehension in case two referential candidates matching the adjective were available in the visual scene. In this case, the adjective would not predict a specific referent. In particular, the question is whether over-specification would still result in a facilitatory effect, in case the visual context supported both a contrastive (Gricean) and a non-contrastive interpretation of the prenominal adjective, i.e., when one of the referential candidates belonged in a contrast pair (cf. the yellow bowl in Fig. 3.1a), while the other was singleton (cf., the watering can in Fig. 3.1a). If so, this should be taken to indicate that the observed facilitation in the OS condition was an effect of over-specification per se. Otherwise, this facilitation is perhaps better explained by the visually-determined surprisal on the noun (which in the OS condition was lower compared to both the MS and the MM conditions). That is, the advantage for over-specification may derive from the fact that it often lowers surprisal.

3.2 Summary

The findings from Experiment 1 demonstrate that ERPs can index a broad spectrum of specificity effects in visually-situated comprehension, and that the N400 component is sensitive to visually-determined surprisal. These findings offer two important insights regarding the online comprehension of referring expressions. Firstly, they shed light on a long-standing debate regarding the influence of over-specification on referential processing. In particular, redundant pre-nominal adjectives were found to facilitate, rather than hinder, comprehension, when the visual contexts were more demanding than just two objects differing in two features (cf. Engelhardt et al., 2011).

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

This is likely because in these visual contexts the pre-nominal adjectives contributed to the reduction of the visually-determined surprisal on the subsequent word (cf. reduced N400 on the noun). It seems, therefore, that referential redundancy per se does not facilitate or hinder comprehension (cf. Gricean approach), but rather its effect depends on the complexity of the situation (e.g., the visual scene, the attributes along which referents differ, etc.). Secondly, the effect of under-specification on online comprehension in visually-situated contexts was found to be different from the effect typically observed in discourse contexts (Nref). Under-specification was found to result in a long-lasting positivity starting around 400 ms after the onset of the ambiguity (the pre-nominal adjective) and continuing through the end of the utterance. This result indicates (a) that listeners are able to rapidly identify unhelpful information, and (b) that referential failure due to under-specification is qualitatively different from referential failure due to an explicit mismatch between the linguistic and visual input, which elicits an increased N400 on the noun.

Chapter 4

The influence of entropy reduction on the comprehension of over-specifications

The previous Chapter addressed a long-standing debate regarding the influence of referential redundancy on visually-situated comprehension. The results from Experiment 1 demonstrated that a redundant pre-nominal adjective may facilitate processing of the upcoming noun (see attenuated N400 on the noun for the OS compared to the MS condition). It is possible, however, that such a facilitation is merely due to the reduced surprisal on the noun in the OS vs. the MS condition: In the OS but not in the MS condition, the adjective selected exactly one object, which was next mentioned by the noun. It is, therefore, not clear whether the facilitation observed for the OS condition was the effect of the visually-determined surprisal on the noun or of over-specification per se. In Chapter 4, we present two experiments – one ERP, the other eye-tracking, using identical stimuli – that aim at disentangling exactly this point. In both experiments, the visual displays that were combined with over-specified instructions included a competitor object fitting Gricean considerations (i.e., it was part of a contrast pair, thus making an adjective necessary), and keeping surprisal similar to the MS conditions. It is, therefore, possible that in these experiments comprehension of over-specified utterances would be hindered, as listeners may expect the contrasting object to be mentioned after the adjective. Such an effect would indicate that listeners attribute a contrastive interpretation to adjectives and bear Gricean expectations in situated communication.

In a visual world eye-tracking study, Sedivy et al. (1999) manipulated exactly this factor using colour and size prenominal adjectives. They found shorter fixation latencies to the target referent when it was part of a contrast pair (minimally-specified reference) compared to when it was not (over-specified reference). The authors interpreted this finding as evidence that participants readily used pragmatic inferencing to inform their interpretation of the utterance as it unfolded. It is, however, possible that this result was due to the specific experimental task rather than due to listeners' preference to interpret the adjective contrastively. Visual scenes (see Fig. 4.1) consisted of four objects: a contrast pair differing in one feature (e.g., a yellow and a pink comb), and two singletons, one bearing the same feature as the target referent (e.g., a yellow bowl) and an unrelated distractor object. While the critical instruction mentioned either of the two referents with the shared feature (i.e., one of the yellow objects), it always followed an instruction that referred to one of the objects in the contrast pair (i.e., one of the two combs). Therefore, an alternative interpretation of the results is possible, namely that participants were faster to fixate the target referent when a contrasting object was available, because their attention was already allocated to the contrast pair and not because the instruction was in this case minimally-specified. Two additional experiments in which the critical instruction came first yielded similar results, but these studies used scalar adjectives such as 'tall', which inherently invoke a comparison between the members of a contrast pair.

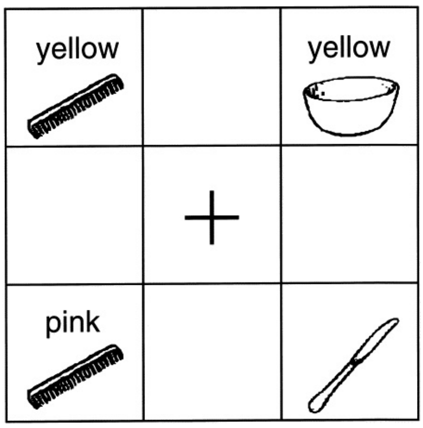


Figure 4.1 Example visual scene from Sedivy et al. (1999).

According to the bounded-rational approach (see Section 2.3), however, it is possible that redundant referring expressions are preferred over minimal descriptions (even if the visual context supports a contrastive interpretation of the adjective, as in Sedivy et al., 1999), because they distribute the same content (the identity of the target referent) across more linguistic units. In other words, over-specifications may facilitate comprehension – Gricean preferences notwithstanding – because they spread the amount of entropy that needs to be reduced for establishing reference (i.e., cognitive effort) across a longer sequence of words. Redundancy may, thus, be efficient, because it helps modulate entropy reduction while also providing addressees with additional cues to guide their visual search for the target referent, making identification faster and less effortful, and ensuring the success of the interaction.

In this Chapter, we set out to examine which of the two approaches better explains listeners' comprehension of over-specifications: the traditional Gricean approach or the bounded-rational approach. We, therefore, address RQ-2 and enquire whether the rate of referential entropy reduction further influences processing, and whether any effect of over-specification would be additive to that of entropy reduction. Furthermore, we revisit RQ-1, asking whether over-specifications are beneficial to comprehension, even when the visual context supports both a contrastive and a non-contrastive interpretation of the adjective – i.e., where two referents match the adjective, one in a contrast pair and one singleton. In Experiments 2 and 3, we, therefore, manipulate *Specificity* and *Entropy Reduction* as orthogonal factors, in order to investigate their independent influence on the online comprehension of referring expressions, and whether their effects would be additive.

In sum, Experiments 2 and 3 investigate the influence of Specificity and Entropy Reduction on visually-situated comprehension. While the instructions always included a prenominal adjective, we manipulated whether the intended referent was a singleton (over-specified reference) or it was part of a contrast set (minimally-specified reference). Thus, we were able to assess whether listeners compute Gricean pragmatic inferences online, and whether their comprehension of the expression is adversely affected when expectations based on those inferences are not met. As in Sedivy et al. (1999), both types of referents (singleton and contrast) were available in the scene regardless of whether instructions were minimally-specified or over-specified. In addition, we examined whether the rate of referential entropy reduction in the expression further influences processing, and whether this influence is additive to any effects of specificity. In Experiment 2, processing effort was assessed by measuring participants' ERPs, while in Experiment 3 we measured their visual attention (log-gaze probabilities). In both experiments we also used the Index of Cognitive Activity (ICA; Marshall, 2000, 2002) as a complementary measure of cognitive effort linking the other two dependent measures, so as to be able and relate the brain responses to the attentional measures (recall that the stimuli tested in the two experiments were identical).

The ICA is a direct measure of cognitive workload that is based on pupillary response. It is well known that fluctuations of the pupil size index cognitive effort in a variety of tasks, including language processing (e.g., Engelhardt et al., 2010; Just and Carpenter, 1993; Scheepers and Crocker, 2004). However, changes in the lighting conditions of the environment are also responsible for pupil dilation. The ICA measures cognitive workload by separating variation in pupil size that is due to cognitive effort or 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 the ICA is particularly suitable for the visual world paradigm since it is robust to the change of fixation position and can thus complement the standard visual attention metrics, in order to assess cognitive effort during linguistic processing. In a related study (Ankener et al., 2018), the ICA was shown to yield reliable results (but see also Sekicki and Staudte, 2018; Vogels et al., 2018, for the use of the ICA in visual world studies).

4.1 Experiment 2

As in Experiment 1, participants in this study attended to spoken instructions to locate a referent in a visual scene, e.g., 'Find the blue ball', combined with displays such as those in Figure 4.2. While the instruction was held constant, scenes differed in whether the intended referent belonged in a contrast set (cf. Fig. 4.2a and b, where a shape competitor is available), or it was a singleton (cf. Fig. 4.2c and d, where there is no shape competitor). Thus, depending on the visual context, the prenominal adjective was either necessary or redundant, and the description was minimally-specified (MS) or over-specified (OS), respectively. In addition to Specificity, we

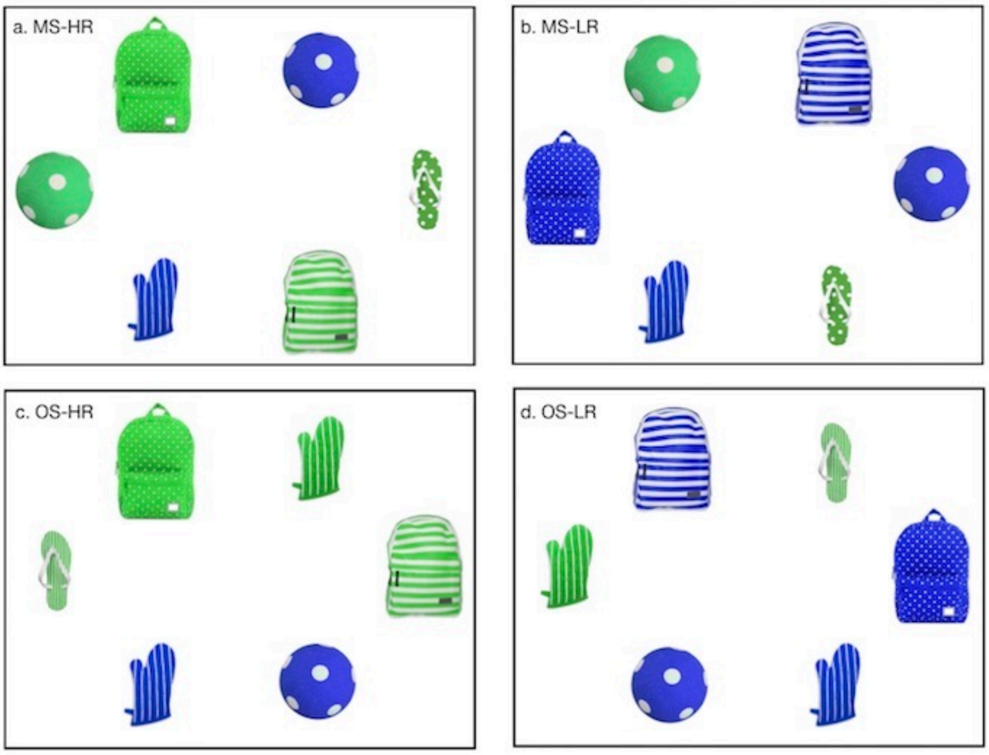


Figure 4.2 Experiment 2. Sample visual scenes for a colour experimental item, paired with the instruction ‘Find the blue ball’.

manipulated the rate of Entropy Reduction due to the adjective, i.e., the number of objects that matched the description at the adjective (cf. two blue objects in Fig. 4.2a and c, and four blue objects in Fig. 4.2b and d). That is, in all conditions before the adjective was heard (i.e., at ‘Find the’), all 6 objects were potential target referents, and entropy was 2.58 bits. The adjective restricted the set of referents to a greater or lesser degree, contributing to a High Reduction (HR) of referential entropy (1.58 bits in Fig. 4.2a and c) or a Low Reduction of referential entropy (0.58 bits in Fig. 4.2b and d), respectively. Importantly, this reduction resulted in a smaller (1 bit) or larger (2 bits) amount of residual entropy, respectively, to be eliminated at the noun. Participants’ EEG and pupil size (based on which the ICA was later calculated) were recorded during the experiment.

We considered two regions for analysis: the adjective, and the noun. Note, however, that in the adjective region only the Entropy Reduction manipulation is of interest, because at this point in the utterance participants were not yet able to determine whether the unfolding expression was minimally-specified or over-specified. Based on the Entropy Reduction Hypothesis (Hale, 2003, 2006), we expected to find effects of processing effort at each reduction point. That is, on the adjective the HR conditions were expected to yield effects associated with increased processing difficulty compared to the LR conditions (e.g., N400). Similarly, we expected ICA on the adjective to be higher in the HR compared to the LR conditions. This pattern was expected to be reversed in the noun region, where residual entropy after the reduction on the adjective was lower in the HR compared to the LR conditions. On the noun, we, therefore, expected the HR conditions to result in effects indexing lower processing effort compared to the LR conditions (e.g., reduced N400, lower ICA values). It is, however, possible that we only observe an effect on the noun, as in Ankener et al. (2018), where a verb that selected for fewer objects did not itself elicit increased ICA values, but did nevertheless result in lower processing effort on the subsequent noun. Regarding Specificity, the Gricean account predicts greater processing effort on the noun (higher ICA values) in the OS compared to the MS condition. By contrast, such a difference is not predicted by the bounded-rational approach, according to which even redundant words may benefit comprehension as long as they reduce referential entropy.

4.1.1 Method

Participants. Forty-four native speakers of German (mean age = 24.2, 19 female), with normal or corrected-to normal vision and no problems with colour perception participated in this experiment. None of the participants had taken part in Experiment 1 or was previously exposed to the stimuli. All participants gave written informed consent prior to the start of the experiment, and were monetarily compensated for their participation. The data from one participant were corrupted and not included in the analyses.

Materials. A new set of visual stimuli was created using 30 of the object pictures from Experiment 1 (for more details, see Appendix A). For ease of counterbalancing, we used three colours (red, blue and green), to match with the number of patterns used (checkered, striped and dotted).

In total, 1320 visual displays were created, of which 960 were used to construct the experimental items, and the rest 360 were used in the fillers. Experimental items were the combination of 4 displays with one spoken instruction (cf. Fig. 4.2). This yielded 240 experimental items, half of which were paired with colour instructions (colour items), and the other half with pattern instructions (pattern items; cf. Fig. B.2 in Appendix B). All experimental displays were created in a way that neither the target feature nor the target referent would be identifiable before hearing the critical words. To this end, six objects were used per display in two colours and two patterns. Two of the objects were singletons, and the rest were paired in two contrast sets, such that the singleton objects could potentially serve as over-specified targets, and the contrast objects could serve as minimally-specified targets, either with colour or with pattern instructions. As in Experiment 1, only same-gender objects appeared in each display, in order to ensure that the determiner would not reveal the target referent, and that the first point of entropy reduction would always be the adjective. Similarly, no phonological competitors were used in the same scene, so that adjective onset would be the first point of disambiguation.

Filler displays differed from experimental displays in several respects. First, 210 filler displays depicted only four objects, introducing some variation in the stimuli set while also making the 6-object experimental displays more complex. Furthermore, half of the filler items were minimally-specified (MS), and the other half were either over-specified (OS) or under-specified (US). In this way, listeners were required to be more attentive (as it was possible that reference would not be resolved), while maintaining a lower proportion of over-specifications, as is normally found in language use (cf. Engelhardt et al., 2006, and references therein). Moreover, all MS and OS filler displays, as well as some of the US ones, contained a set of three same-shape objects (e.g., three balls) differing for both colour and pattern, thus making the use of a second adjective necessary for disambiguation. The rest of the US fillers were similar in structure to the experimental displays, but failed to establish reference (e.g., 'the green rucksack' when two objects fit the description in Fig. 4.2). Twelve fillers were used as practice items in a familiarisation session before the experiment.

Experimental displays were paired with spoken instructions containing a prenominally modified referring expression like 'Find the blue ball' in German ('Finde den blauen Ball'), while filler instructions could mention one, two or no modifiers. The order of mention of colour and pattern adjectives was counterbalanced in the two-modifier fillers. Audio stimuli were the same as in Experiment 1. Mean word duration was 397.2 ms (*SD* = 49.6) for colour adjectives, 605.1 (*SD* = 75.1) for pattern adjectives, and 557.2 ms (*SD*=75.7) for the nouns.

Stimuli were divided into 4 lists of 588 trials, so that one version of an item was in each list, and no participants saw more than one condition of a given item. Lists were pseudo-randomised for each participant, making sure that 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. The experiment was implemented and run using E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Procedure. After participants read the instructions of the experiment, they were seated at a distance of approximately 60 cm in front of the monitor. As in Experiment 1, participants were first presented with displays of each object in all colours and patterns accompanied by pre-recorded audios of each object's type. A familiarisation phase was then administered, during which the experimenter gave feedback after each trial, to make sure that the task was clear to the participant before the experiment began. Each experimental session was divided into 7 blocks, in between which participants could take short breaks. One experimental session took on average 90 minutes to complete.

Visual stimuli were presented at a resolution of 1680 x 1050 pixels. A trial started with a 2.5 seconds preview of the visual scene (see Fig. 4.3). After this time a cross appeared in the middle of the screen, and 500 ms later the spoken instruction was played. At the offset of the audio the cross disappeared, while the objects remained on the screen for another 500 ms. Participants were asked to fixate the cross while it was on the screen. At the end of each trial, participants saw a display (see Fig. C.1 in Appendix C) prompting them to indicate which side of the screen the target object appeared on. Responses were given in the form of button presses in a Cedrus response pad (Cedrus Corporation, San Pedro, California, USA).

The EEG was recorded from 26 Ag/AgCl electrodes placed on the scalp according to the standard 10-20 system. The signal was amplified by a BrainAmps DC amplifier (Brain Products, GmbH, Munich, Germany), and digitised at a sampling rate of 500 Hz The EEG was re-referenced offline to the average of both mastoid electrodes. Eye

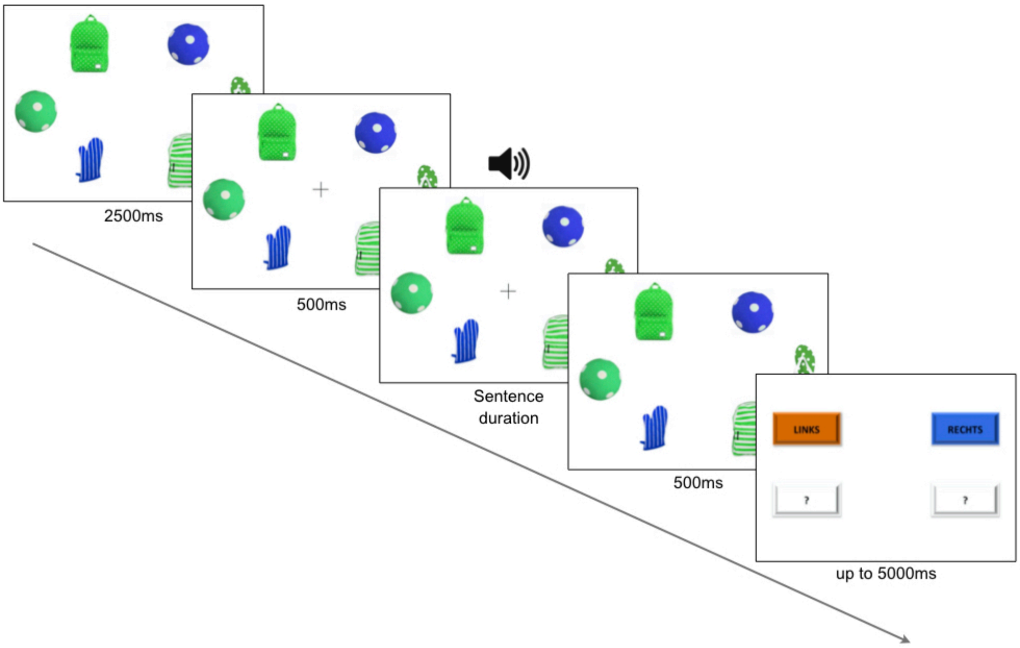


Figure 4.3 Experiment 2. Example trial sequence and timing of events.

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 5 kΩ.

In order to measure the ICA, apart from the EEG we also recorded participants' pupillary activity. For this reason, we used an SMI RED 250 eye-tracker (SensoMotoric Instruments GmbH, Berlin, Germany) attached to the bottom of the 22-inch Dell stimulus presentation monitor. The sampling rate was 250 Hz. Calibration of the eye-tracker was performed at the beginning of each block.

Data analysis. The offline processing of the EEG data was performed using BrainVision Analyzer 2 (Brain Products, GmbH, Munich, Germany). The EEG signal was filtered (30 Hz high cut-off, and 0.1 Hz low cut-off) and re-referenced offline to the average of the two mastoid electrodes. Single-participant averages were then computed in a 1000 ms window per condition relative to the onset of the adjective (e.g., 'blaue') and head noun (e.g., 'Ball'), and aligned to a 200 ms pre-stimulus baseline. Trials were semi-automatically screened offline for eye movements, blinks, electrode drifts, and amplifier blocking. After artefact rejection, 10 participants with less than 30 trials per condition were excluded from the analyses. Only artefact-free ERP averages time-locked to the onset of the critical regions entered the analyses.

To calculate the ICA we used the BeGaze software equipped with the ICA Module (SensoMotoric Instruments, GmbH, Berlin, Germany) and Workload RT (EyeTracking, Inc., Solana Beach, CA, USA). Since the ICA values output by the BeGaze software are too coarse-grained for the type of effects we expected, we used the ICA Coefficients to compute ICA values per 100 ms (see Demberg and Sayeed, 2016, for more details). Data points with a pupil diameter smaller than 2.5 SD per participant were eliminated, and a mean ICA value for both eyes was calculated. We compared mean ICA values across conditions within a window of 600 ms starting from the middle of each region (cf. Sekicki and Staudte, 2018).

We analysed participants' ERPs, ICA in two time-windows, after adjective and after noun onset, as well as their response times (RTs). For the ERPs, we followed the same procedure as in Experiment 1. We performed ANOVAs (ez package, version 4.4.0; Lawrence, 2011) in R (version 3.5.1; R Core Team, 2018) at a Midline column containing electrodes Fz, Cz and Pz, and two lateral columns: the Medial column, containing electrodes FC1, C3, CP1, FC2, C4 and CP2, and the Lateral column, containing electrodes F3, FC5, CP5, P3, F4, FC6, CP6 and P4. For the Medial and Lateral columns, within-subjects factors were Entropy Reduction (levels: HR, LR), Specificity (levels: MS, OS), Longitude (levels: anterior, central, posterior), and Laterality (left, right). For the Midline column, only the factors of Entropy Reduction, Specificity and Longitude were included in the ANOVAs. We further included the Feature of the target referent (levels: Colour, Pattern) as a fixed factor in the ANOVAs. Where appropriate, we will report the Greenhouse-Geisser corrected p-value (Greenhouse and Geisser, 1959) with the original degrees of freedom.

For the analyses of the ICA and of RTs, we fitted (generalised) linear mixed models (lme4 package; Bates et al., 2015) in R (version 3.5.1; R Core Team, 2018) including Entropy Reduction, Specificity and Feature as fixed factors, and crossed random intercepts and slopes for participants and items. All factors were contrast-coded, with positive contrast-coding (0.5) for the levels of HR, MS and Colour, and negative contrast-coding (–0.5) for LR, OS and Pattern. Whenever the maximal models did not converge, we simplified the random effects structure as suggested by Barr et al. (2013).

As discussed above, the Specificity manipulation is not relevant in the adjective region, as specificity of the referring expression is determined by the noun. Therefore, for the ERP and ICA analyses in the adjective region, we collapsed across the MS and OS conditions, and included only Entropy Reduction and Feature as fixed factors in the models. In the noun region, Specificity was also included in the analyses.

4.1.2 Results

Response accuracy was higher than 90% in all conditions for most participants. The data from one participant with accuracy less than 75% were excluded from further analyses. The rest of the participants gave the correct answer at a rate of 93.3% in the MS-HR condition, 92.3% in the MS-LR condition, 96.3% in the OS-HR condition, and 95.8% in the OS-LR condition. We analysed response accuracy using generalised linear mixed models (lme4 package Bates et al., 2015) in R (version 3.5.1; R Core Team, 2018) with a binomial function, including Entropy Reduction, Specificity and Feature as fixed factors, and crossed random intercepts for participants and items. Results showed two main effects, of Specificity (β = -0.703, SE = 0.11, z = -6.419, $p < .001$) and of Feature (β = 1.174, SE = 0.172, z = 6.816, $p < .001$), such that the OS conditions as well as the Colour items resulted in higher response accuracy. All analyses described below included only trials with correct responses.

Response times. As in Experiment 1, RTs were time-locked to the onset of the prompt display, and analyses were carried out on log-transformed RTs. As is seen in Figure 4.4, participants responded faster in the HR (473 ms, SD = 368) compared to the LR conditions (514 ms, SD = 414; β = -0.077, SE = 0.017, t = -4.483, $p < .001$), and faster in the OS (465 ms, SD = 347) compared to the MS conditions (522 ms, SD = 432; β = 0.086, SE = 0.018, t = 4.644, $p < .001$). Faster responses were also found with colour (419 ms, SD = 302) compared to pattern adjectives (568 ms, SD = 454; β = -0.245, SE = 0.036, t = -6.847, $p < .001$). In both Colour and Pattern items, Entropy Reduction and Specificity significantly affected response times. In Colour items, participants responded faster in the HR (399 ms, SD = 271) compared to LR conditions (439 ms, SD = 328; β = -0.086, SE = 0.019, t = -4.515, $p < .001$), as well as in the OS (404 ms, SD = 264) compared to the MS conditions (434 ms, SD = 334; β = 0.044, SE = 0.019, t = 2.296, p = .023). Similarly, in Pattern items participants' responses were faster in the HR (547 ms, SD = 433) compared to the LR conditions

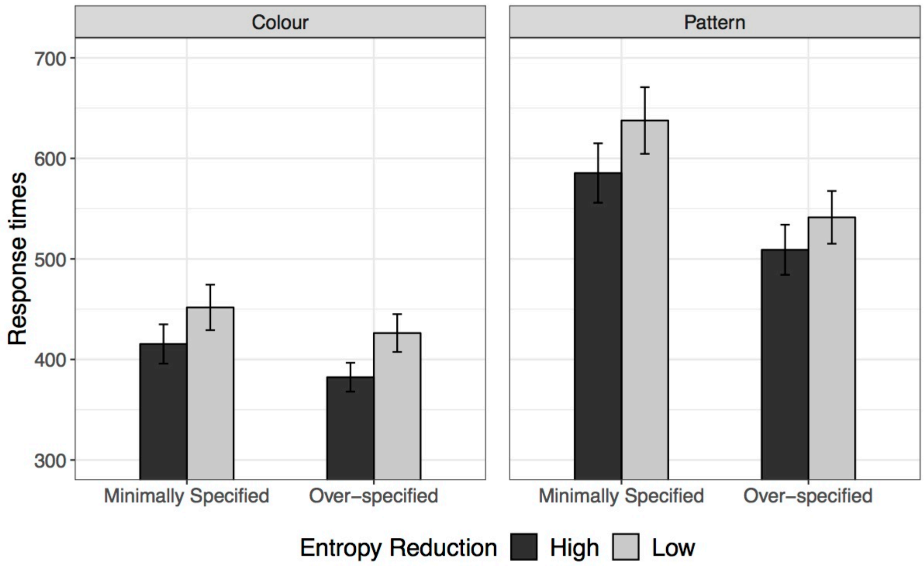


Figure 4.4 Experiment 2. Response times per condition in Colour and Pattern items. Error bars represent 95% CIs.

(589 ms, *SD* = 473; β = -0.07, *SE* = 0.026, *t* = -2.664, *p* = .012), as well as in the OS (525 ms, *SD* = 404) compared to the MS conditions (611 ms, *SD* = 496; β = 0.128, *SE* = 0.032, *t* = 3.959, *p* < .001).

Index of Cognitive Activity. None of the experimental manipulations was found to significantly influence ICA values (*p* > .05), neither in the adjective nor in the noun region (see Fig. 4.5). In the adjective region more specifically, both Entropy Reduction and Feature resulted in non-significant effects (*p* > .05). In the noun region as well, none of the manipulated factors reached significance (*p* > .05).

ERPs. In the adjective region, we collapsed across Specificity, and only Entropy Reduction and Feature were considered for analysis (see Data analysis in Section 4.1.1). Visual inspection of the ERP waveforms time-locked to the onset of the adjective did not reveal differences between the Entropy Reduction conditions, but indicated a difference between Colour and Pattern items. Specifically, the adjective yielded a larger positivity starting around 400 ms in Pattern items compared to Colour items.

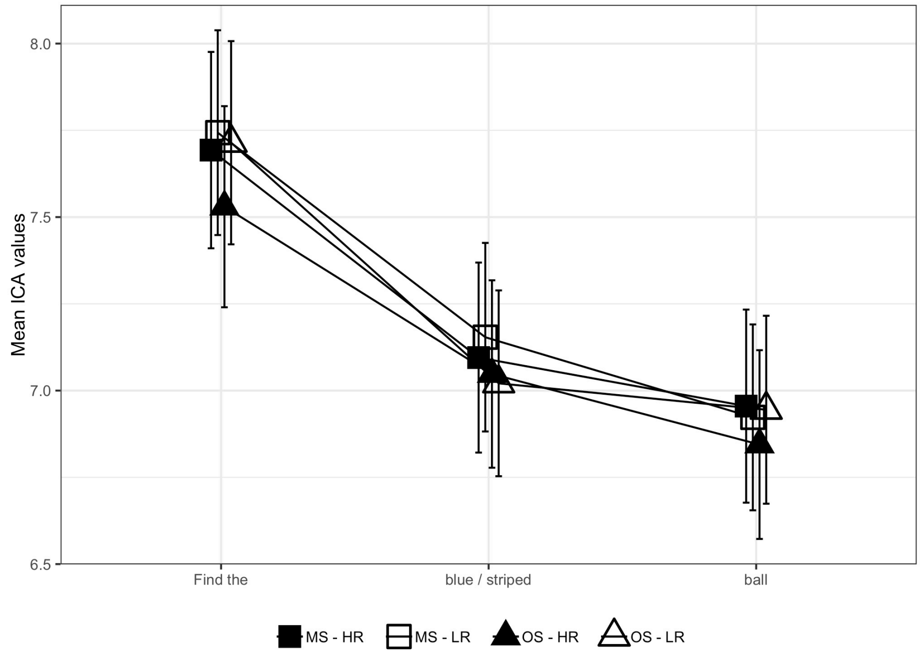


Figure 4.5 Experiment 2. Mean ICA values per condition in each region. Error bars represent 95% CIs.

As in Experiment 1, we performed analyses in a time-window between 400 and 600 ms (see shaded area in Fig. 4.6, because adjectives were on average 500 ms long (397 ms in colour, and 605 ms in pattern items), so any effects later than that should be associated with processing of the subsequent noun. For Entropy Reduction, the ANOVAs showed no significant differences between the HR and LR conditions in any of the recording sites ($p > .05$). On the other hand, Feature yielded a main effect in all midline ($F(1,31) = 11.73$, $p = .002$, $\eta_p^2 = .054$), medial ($F(1,31) = 13.61$, p < .001, $\eta_p^2 = .064$) and lateral (F(1,31) = 16.73, p < .001, $\eta_p^2 = .058$) electrodes, with a larger positivity for Pattern compared to Colour items. Moreover, the Feature x Longitude x Laterality interaction was marginally significant ($F(2,62) = 2.83$, $p = .07$, $\eta_p^2 < .001$) in the medial sites, indicating that the positivity observed for Pattern items was larger over centro-parietal electrode sites. We followed up the effect of Feature with separate analyses for Colour and Pattern items, and we found no significant effects of Entropy Reduction in any of the electrode columns ($p > .05$). Bearing in mind that the difference in the duration of colour and pattern adjectives is roughly 200 ms on average, and given that in both Colour and Pattern items the subsequent noun

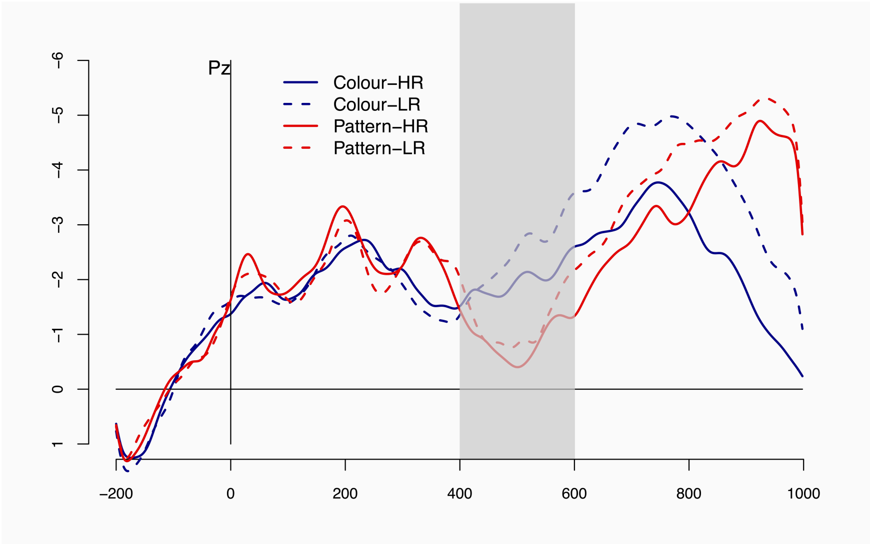


Figure 4.6 Experiment 2. Averaged ERPs at electrode Pz time-locked to adjective onset. The shaded area indicates the 400-600 ms analysis window. Negative voltages are plotted upward.

results in a negativity (see Fig. 4.8 and Fig. 4.11, respectively) – which, however, starts around 200 ms later in Pattern items – it seems that the effect of Feature should likely be attributed to the longer duration of pattern adjectives. That is, because pattern adjectives are on average 200 ms longer than colour adjectives, the negativity observed in the subsequent time-window, appears as a positivity in the time-window between 400 and 600 ms post adjective onset.

In the noun region, visual inspection of the data indicated a larger negativity peaking around 400 ms after noun onset for the LR compared to the HR conditions (see Fig. 4.7). In this region, Specificity was also included as a factor in the analyses. The OS conditions resulted in a larger negativity compared to the MS conditions peaking at 400 ms after noun onset. This effect was followed by a larger positivity for the OS compared to the MS conditions, which started at around 600 ms after the onset of the noun and continued through the end of the noun region. Table 4.1 summarises the results in the two time-windows.

Separate analyses were performed for Colour and for Pattern items. For Colour items (see Fig. 4.8), in the N400 time-window we found a broadly distributed effect of Entropy Reduction, which was significant in all midline ($F(1,31) = 7.913$, $p = .008$,

Table 4.1 Experiment 2. ANOVAs on ERPs in the noun region, in the N400 (300-500 ms) and the P600 (600-900 ms) time-windows

	Effect	N400 time-window			P600 time-window		
		F(df)	p	η_p^2	F(df)	p	η_p^2
Midline	Red	6.672(1,31)	.014*	.18	0.694(1,31)	.411	.02
	Spec	9.094(1,31)	.005*	.23	0.729(1,31)	.399	.02
	Feat	16.308(1,31)	<.001*	.34	36.577(1,31)	<.001*	.54
	Spec:Long	0.195(2,62)	.823	.00	5.607(2,62)	<.001*	.31
	Feat:Long	13.275(2,62)	<.001*	.30	5.607(2,62)	.020*	.15
Medial	Red	8.040(1,31)	.008*	.20	1.62(1,31)	.212	.05
	Spec	10.357(1,31)	.003*	.25	0.855(1,31)	.362	.03
	Feat	19.761(1,31)	<.001*	.39	45.424(1,31)	< .001 *	.59
	Red:Long	7.282(2,62)	.004*	.19	2.09(2,62)	.145	.06
	Spec:Long	0.054(2,62)	.908	.00	7.871(2,62)	.001*	.20
	Feat:Long	12.092(2,62)	<.001*	.28	2.395(2,62)	.108	.07
Lateral	Red	7.804(1,31)	.009*	.20	0.695(1,31)	.411	.02
	Spec	12.033(1,31)	.001*	.28	0.121(1,31)	.730	.00
	Feat	18.9(1,31)	<.001*	.38	40.497(1,31)	<.001*	.57
	Red:Long	4.233(2,62)	.048*	.12	0.044(2,62)	.835	.00
	Spec:Long	0.004(2,62)	.947	.00	19.297(2,62)	<.001*	.38
	Feat:Long	19.591(2,62)	<.001*	.39	3.128(2,62)	.087	.09
	Red:Long:Lat	5.193(1,31)	.029*	.14	2.434(1,31)	.128	.07
	Red:Spec:Feat	1.739(1,31)	.197	.05	6.462(1,31)	.016*	.17

Note: Red = Entropy Reduction, Spec = Specificity, Feat = Feature, Long = Longitude, Lat = Laterality

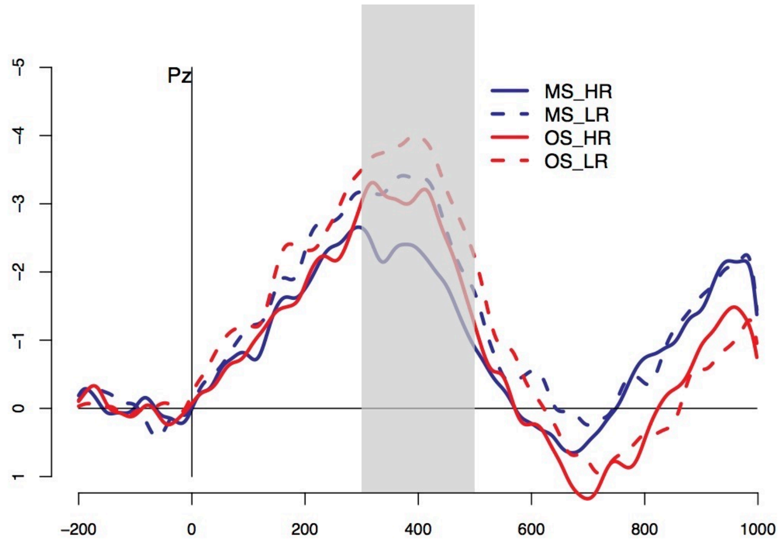


Figure 4.7 Experiment 2. Averaged ERPs at electrode Pz, time-locked to noun onset. The shaded area indicates the 300-500 ms analysis window. Negative voltages are plotted upward.

$\eta_p^2 = .2$), medial ($F(1,31) = 8.353$, $p = .007$, $\eta_p^2 = .212$), and lateral ($F(1,31) = 10.311$, $p = .003$, $\eta_p^2 = .25$) channels, with the LR conditions being more negative compared to the HR conditions (see in Fig. 4.9a). We found no effect of Specificity ($p > .05$) in any of the recording sites, indicating that noun retrieval did not differ between the MS and OS conditions (see Fig. 4.9b). However, the Specificity x Longitude x Laterality interaction ($F(1,31) = 4.375$, $p = .045$, $\eta_p^2 = .12$) was significant in the lateral channels, indexing a negativity for the OS compared to the MS conditions over right anterior electrode sites.

In the P600 time-window (600-900 ms post noun onset), we found no effects ($p > .05$) of Entropy Reduction in any of the recording sites (see Fig. 4.10a, where the negativity from the previous time-window is sustained in the current one). There was, however, a significant Specificity x Longitude interaction in all midline ($F(2,62) = 6.916$, $p = .002$, $\eta_p^2 = .18$), medial ($F(2,62) = 4.39$, $p = .016$, $\eta_p^2 = .12$) and lateral $F(2,62) = 9.796$, $p = .004$, $\eta_p^2 = .24$) sites, with a larger positivity evoked for the OS compared to the MS conditions over centro-posterior electrodes (see Fig. 4.10b). Finally, we also observed a significant interaction between Entropy Reduction and Specificity over lateral electrode sites ($F(2,62) = 4.265$, $p = .047$, $\eta_p^2 = .12$), with the OS-HR condition

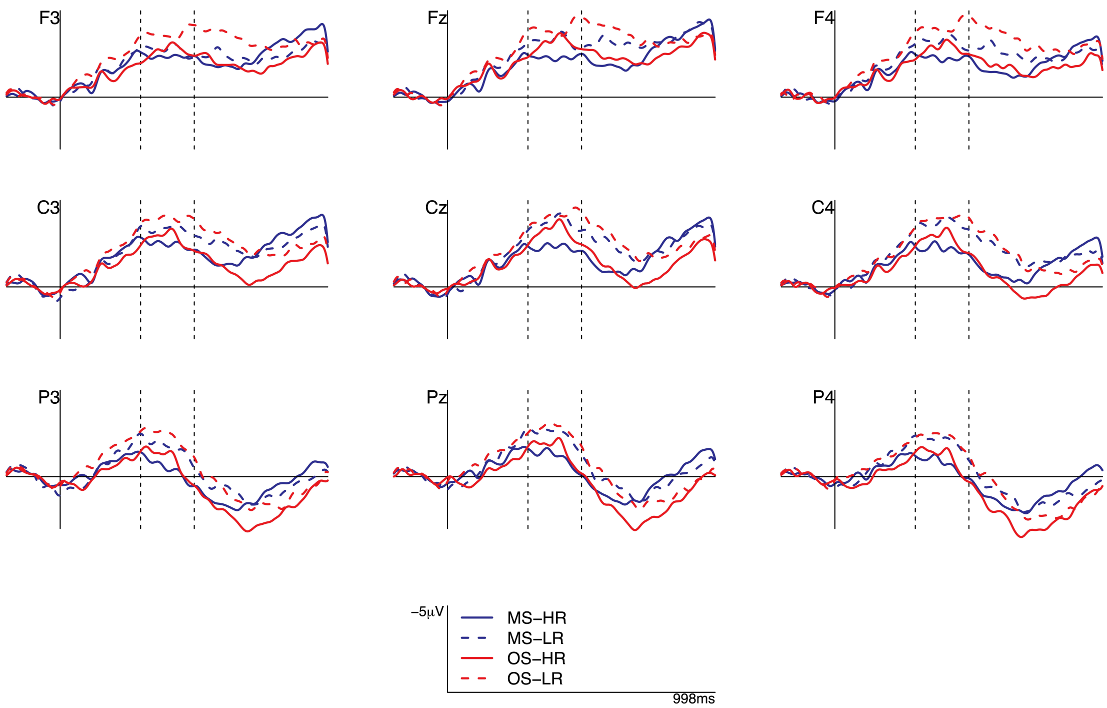
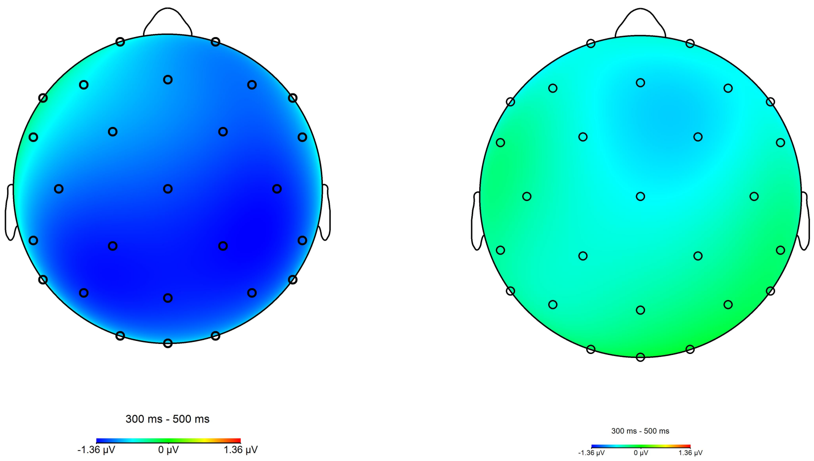


Figure 4.8 Experiment 2. Averaged ERPs for Colour items, time-locked to noun onset. The dotted lines indicate the 300-500 ms analysis window. Negative voltages are plotted upward.

yielding a larger positivity compared to the other three conditions (see red solid line in Fig. 4.8).

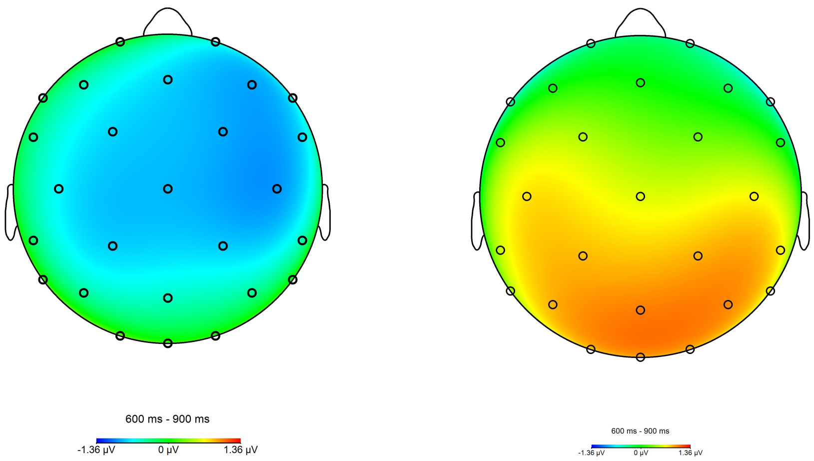
For Pattern items (see Fig. 4.11), we found no main effect of Entropy Reduction in the N400 time-window ($p > .05$). We did, however, observe a significant Entropy Reduction x Longitude interaction in the medial sites ($F(2,62) = 4.713$, $p = .012$, $\eta_p^2 = .13$), with the LR condition yielding a larger negativity compared to the HR condition over centro-parietal electrodes (see Fig. 4.12a).

Regarding Specificity, we found a main effect in all midline ($F(1,31) = 8.768$, $p = .006$, $\eta_p^2 = .22$), medial ($F(1,31) = 10.346$, $p = .003$, $\eta_p^2 = .25$) and lateral ($F(1,31) = 12.43$, $p = .001$, $\eta_p^2 = .29$) sites, with the N400 amplitude being more negative in the OS conditions than in the MS conditions. Furthermore, we also found a Specificity x Laterality interaction in both the medial ($F(1,31) = 5.546$, $p = .025$, $\eta_p^2 = .15$) and lateral ($F(1,31) = 5.394$, $p = .026$, $\eta_p^2 = .15$) electrode sites, indicating that the N400 for



(a) LR minus HR (b) OS minus MS

Figure 4.9 Experiment 2. Topographic maps for Colour items in the time-window 300-500 ms post noun onset



(a) LR minus HR (b) OS minus MS

Figure 4.10 Experiment 2. Topographic maps for Colour items in the time-window 600-900 ms post noun onset

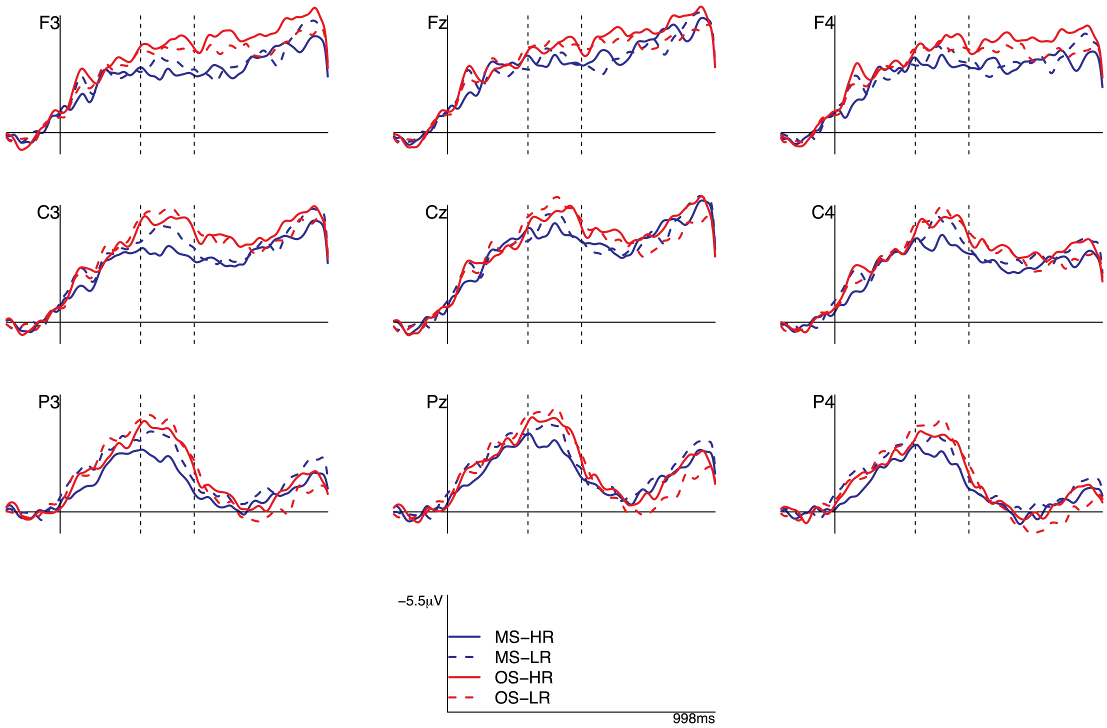
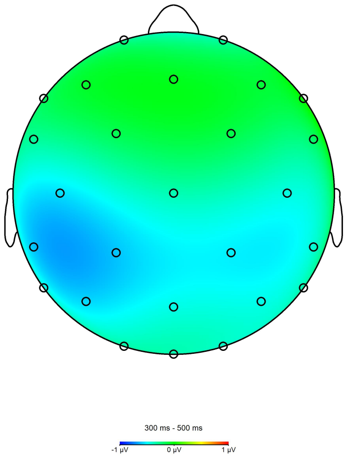


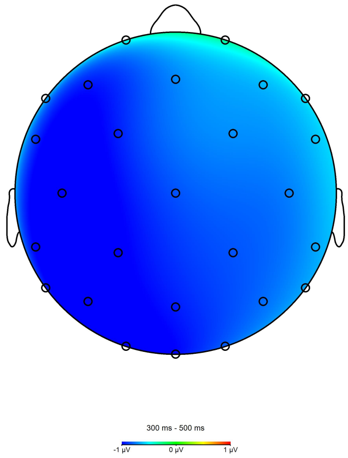
Figure 4.11 Experiment 2. Averaged ERPs for Pattern items time-locked to noun onset. The dotted lines indicate the 300-500 ms analysis window. Negative voltages are plotted upward.

the OS vs. MS conditions was more pronounced over left centro-parietal recording sites (see Fig. 4.12b).

In the P600 time-window, we found no effects of Entropy Reduction (see Fig. 4.13a). As for Specificity, we observed a Specificity x Longitude interaction in all midline (F(2,62) = 3.467, p = .037, η_p^2 = .1), medial (F(2,62) = 4.389, p = .016, η_p^2 = .12), and lateral (F(1,62) = 6.306, p = .017, η_p^2 = .17) sites, with a larger positivity evoked in the OS compared to the MS conditions. This effect was small, but can be seen more clearly if we move the time-window by 100 ms, i.e., in the 700-1000 ms time-window (see Fig. 4.13b). The observed positivity was, therefore, sustained.

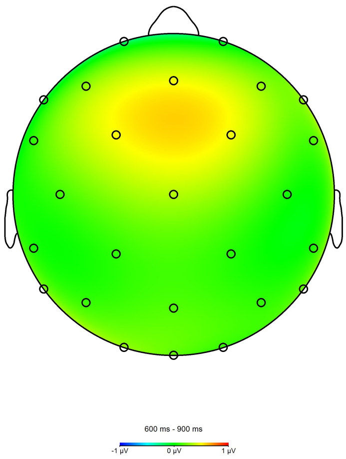


(a) LR minus HR

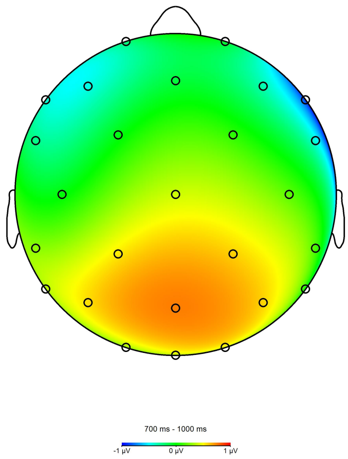


(b) OS minus MS

Figure 4.12 Experiment 2. Topographic maps for Pattern items in the time-window 300-500 ms post noun onset.



(a) LR minus HR



(b) OS minus MS

Figure 4.13 Experiment 2. Topographic maps for Pattern items in the time-window 600-900 ms for LR vs. HR and 700-1000 ms for OS vs. MS post noun onset

4.1.3 Discussion

Our goal in Experiment 2 was twofold: (a) to address RQ-2, asking whether the rate at which referential entropy is reduced during the expression influences comprehension, and (b) to re-visit RQ-1 and establish whether the comprehension of over-specifications is hindered or facilitated relative to that of minimal descriptions, and whether any effect of over-specification is additive to that of entropy reduction. We recorded participants' brain responses and ICA as they attended to spoken instructions to locate a target referent in a visual scene presented on a screen in front of them. The instructions always contained a prenominal (colour or pattern) adjective and a head noun, and were minimally-specified when the target referent was part of a contrast pair (i.e., the adjective was necessary to identify the target), and over-specified when the target referent was singleton (i.e., the adjective was redundant). Moreover, the adjective could select either two or four objects carrying the same property, which were disambiguated by the noun. The adjective, therefore, reduced referential entropy at a higher rate (when it selected two objects) or at a lower rate (when it selected four objects).

Our results draw an intricate picture. First, the ERP results indicate that the processing of the noun in over-specified expressions differs depending on the nature of the redundant prenominal adjective. More specifically, in Pattern items the adjective resulted in a larger N400 effect on the subsequent noun for the OS compared to the MS conditions. This effect indicates that the retrieval of the noun was more cumbersome when the preceding pattern adjective was redundant compared to when it was necessary. In Colour items, no such effect was observed; the waveforms for the OS and MS conditions patterned together in the N400 time-window. On the other hand, in both Colour and Pattern items the comparison between the OS and MS conditions elicited a P600 effect. This effect was stronger in Colour than in Pattern items, while in Colour items it was more enhanced for the HR condition (OS-HR interaction). Recall that in the OS-HR condition it was easier to identify the target referent compared to the MS-HR condition, where an object of the same type as the target (cf. the green ball in Fig. 4.2a) appeared on the other side of the visual display, possibly causing interference in the target identification process. We interpret this P600 as evidence of a task-related effect, essentially as an instance of the P300, where the enhanced positivity is associated with ease in performing the task (cf. Fabiani et al., 1987; Magliero et al., 1984; Sassenhagen et al., 2014). This interpretation is corroborated by the RTs and accuracy results, which show that, despite any processing cost evoked by redundant pattern adjectives, over-specifications result in faster and more accurate responses than minimal descriptions across the board. We can now disentangle between the two possible interpretations of the P600 found for the US vs. MS conditions in Experiment 1. Recall that, even though in both conditions the adjective ('yellow') selected two objects, only in the US condition it was possible to know the type of the target referent already on the adjective, because in that condition both objects selected by the adjective were of the same type (cf. the two bowls in Fig. 3.1c). It seems, therefore, that the P600 effect for the US condition reflects participants' realisation that the adjective picks two referents of the same type, and the ensuing readiness to push the appropriate button (the question-mark button), in case a second adjective disambiguating between the two objects did not come in.

Secondly, we found evidence that the rate at which referential entropy is reduced across the utterance influences situated comprehension. While no effects of Entropy Reduction were observed in the adjective region, Entropy Reduction was found to modulate the N400 on the noun, with the LR condition eliciting a larger negativity compared to the HR condition, in line with the Entropy Reduction Hypothesis (Hale, 2003, 2006). As explained above, residual entropy on the noun was larger in the LR conditions following a small reduction of entropy on the adjective, compared to the HR conditions, where entropy reduction on the adjective was higher. Therefore, a larger amount of entropy was left to be reduced on the noun in the LR conditions, which resulted in difficulty with lexical retrieval on the noun compared to the HR conditions (N400 effect). This effect was broadly distributed after colour adjectives, and localised over centro-parietal sites after pattern adjectives. Similarly, a high reduction of entropy on the adjective enhanced participants' RTs.

One interesting outcome is that the ICA was not found to be affected by any of the factors manipulated in the experiment. ICA values were equally high for both Entropy Reduction conditions and Features in the adjective time-window. In the noun time-window as well, we found no significant effects. One possibility is that the ICA is not sensitive to the kinds of manipulations that we used. It is, however, also possible that the absence of an effect on the ICA is rather a by-product of simultaneously recording the EEG, which required that participants make no eye movements while listening to the instructions. Experiment 3, which employs only eye-tracking, addresses this issue.

In sum, Experiment 2 examined the processing of over-specified referring expressions, and the role of referential entropy reduction in visually-situated communication. Results showed that over-specifications are processed differently depending on whether the redundant adjective was a colour or pattern adjective. While noun retrieval was more cumbersome after a redundant pattern adjective compared to a necessary pattern adjective (N400 for OS), no such effect was found with colour adjectives. Additionally, the rate of referential entropy reduction was also found to influence processing: A high reduction of entropy on the adjective resulted in facilitated retrieval on the noun. Even though this effect was found with both colour and pattern adjectives, it was stronger in Colour items. Surprisingly, ICA values were equally high in all conditions and across Colour and Pattern items, suggesting that the cognitive load that listeners experienced was similar across conditions, and colour and pattern features.

4.2 Experiment 3

In Experiment 3 we aimed to replicate the results from Experiment 2 using eye-tracking and ICA, that is without recording participants' EEG.[1] We recorded participants' pupillary responses and eye movements as they attended to stimuli identical to those in Experiment 2 (repeated in Fig. 4.14 for convenience). In this Experiment, we considered the ICA as a measure of comprehension difficulty, fixation probabilities as a measure of visual attention, and response times for comparison to the previous experiment.

As in Experiment 2, we examined two regions of interest: the adjective, and the noun. In the adjective region, only the Entropy Reduction manipulation was relevant; it was not possible for participants to determine whether the unfolding expression was minimally-specified or over-specified on the adjective. In case the ICA is sensitive to our manipulation, we expected Entropy Reduction to modulate the ICA, with higher reduction rates resulting in increased processing difficulty, according to the Entropy Reduction Hypothesis (Hale, 2003, 2006). More specifically, ICA values on the adjective should be higher in the HR compared to the LR conditions. In contrast, ICA values on the noun should be lower in the HR compared to the LR conditions, since residual entropy on the noun in HR conditions is lower due to the

[1]This study was published in a journal article (see Tourtouri et al., 2019).

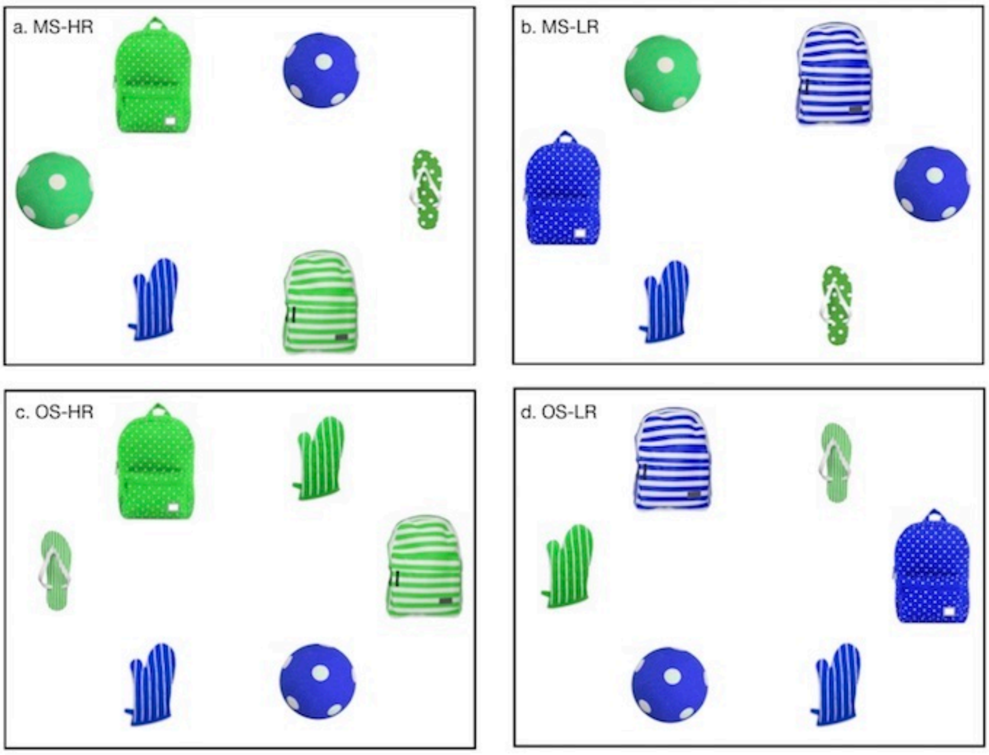


Figure 4.14 Experiment 3. Sample visual scenes for a colour experimental item, paired with the instruction 'Find the blue ball'.

high reduction rate on the preceding adjective. It is, however, possible that we only observe an effect on the noun, as in Ankener et al. (2018), where a verb that selected for fewer objects did not itself elicit increased ICA values, but did, nevertheless, result in lower processing effort on the subsequent noun. On the other hand, in case the ICA is not sensitive to our manipulation, we should see no effects of Entropy Reduction, as in Experiment 2.

Even though it was not possible to determine whether the adjective was necessary (MS conditions) or redundant (OS conditions) before hearing the noun, anticipatory eye movements triggered by the adjective might reveal how listeners interpret the prenominal adjective (cf. Weber et al., 2006), as displays always contained one object that fitted a contrastive reading of the adjective (cf. the blue ball in Fig. 4.14a and b, and the blue mitt in Fig. 4.14c and d), and one singleton object that did not match a contrastive reading (cf. the mitt in Fig. 4.14a and b, and the ball in Fig. 4.14c and d). If listeners are Gricean (i.e., if they assume that adjectives should be used only in relation to a contrast between objects, rather than simply provide redundant information), then the adjective should trigger more anticipatory eye movements towards the contrast object compared to the singleton object (cf. Sedivy et al., 1999). Based on the results of Experiment 2, we may expect that such effects suggestive of Gricean processing should manifest only for pattern items and not for colour items (recall that the N400 effect for the OS vs. MS conditions was observed only for pattern items). If, on the other hand, listeners do not adhere to Gricean principles, but just utilise any information in the utterance to reduce uncertainty regarding the target referent, looks to the contrast and singleton objects should be similar.

In the noun region, the Gricean account predicts greater processing effort (higher ICA values) in the OS conditions compared to the MS conditions. By contrast, the bounded-rational account, according to which redundancy may be preferred because it distributes information (i.e., processing effort) across a longer sequence of words, does not predict such a difference. Moreover, if redundant prenominal adjectives facilitate processing by reducing referential entropy, this should be manifest in an interaction between Specificity and Entropy Reduction, with a larger benefit (lower ICA values) in the OS-HR condition (cf. Fig. 4.14c; recall that in Exp. 2 this condition elicited an increased P600 in Colour items). As visual attention (proportion of fixations) is primarily informative regarding expectations of upcoming material, we do not expect it to reveal anything beyond correct identification of the target on the noun, i.e., more fixations towards the target vs. the competitor object. Note that the *target* object is part of the contrast pair in the MS conditions, but it is singleton in the OS conditions (cf. the blue ball in Fig. 4.14); the *competitor* object is the other referent matching the description up to the adjective, and it is singleton in the MS conditions, but it belongs in a contrast pair in the OS conditions (cf. the blue mitt in Fig. 4.14).

4.2.1 Method

Participants. Twenty-four native speakers of German (mean age = 25, 17 female), with normal or corrected-to normal vision and no problems with colour perception were recruited through the Saarland University Psycholinguistics Group's participant database. None of them was previously exposed to the stimuli. Participants gave written informed consent prior to their participation, and were monetarily compensated.

Materials. Experiment 3 used half of the experimental materials of Experiment 2. It was taken care, however, that all objects appeared as targets an equal number of times, and that target colour, pattern and position were counterbalanced across stimuli. Overall, 660 visual displays were used, 480 of which were used to construct 120 experimental items, and the rest were used for the fillers.

Stimuli were divided into 4 lists of 288 trials so that one version of an item was in each list, and no participants saw more than one condition of a given item. Lists were pseudo-randomised for each participant, making sure that 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. The experiment was implemented and run using E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Procedure. Participants' eye movements were tracked at a rate of 250 Hz using an SMI RED 250 eye tracker (SensoMotoric Instruments GmbH, Berlin, Germany) attached to the bottom of the 22-inch stimulus presentation Dell monitor. Participants read the instructions of the experiment, and were seated at a distance of approximately 60 cm in front of the monitor. A chinrest was used to minimise head movements. A familiarisation phase was first administered, during which the experimenter gave feedback after each trial, to ensure that the task was clear to the participant before the experiment began. Each experimental session was divided into 4 blocks, in between which participants could take short breaks. Calibration was performed at the beginning of each block. On average, participants needed 40 minutes to complete the experiment.

Visual stimuli were presented at a resolution of 1680 x 1050 pixels. At the beginning of each trial a cross appeared in the middle of the display for a period controlled by the experimenter (see Fig. 4.15). After that, the objects appeared while the cross remained on the screen for another 500 ms. The audio instruction was played 1500 ms later. After the end of the instruction, the objects remained on the screen for a wrap-up period of 500 ms. At the end of the trial, participants were prompted to indicate which side of the screen the target referent was on, or whether it was not possible to tell (under-specified fillers) by pressing the corresponding button on a

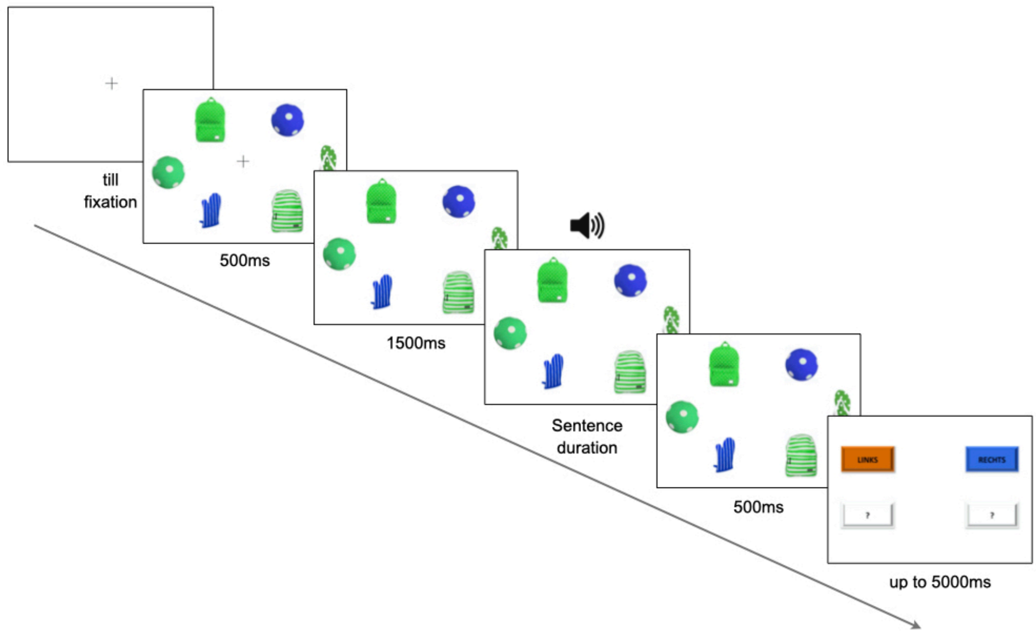


Figure 4.15 Experiment 3. Example trial sequence and timing of events.

Cedrus response pad (Cedrus Corporation, San Pedro, California, USA) placed in front of them.

Data analysis. We analysed the ICA and gaze probabilities in two time-windows, after adjective and after noun onset, as well as response times. For all analyses, we fitted (generalised) linear mixed models (lme4 package; Bates et al., 2015) in R (version 3.5.1; R Core Team 2018) including *Entropy Reduction* (HR vs. LR) and *Specificity* (MS vs. OS) as well as the *Feature* of the target referent (Colour vs. Pattern) as fixed factors, and crossed random intercepts and slopes for participants and items. All factors were contrast coded, with positive contrast coding (0.5) for the levels of HR, MS and Colour, and negative contrast coding (-0.5) for LR, OS and Pattern. Whenever the maximal models did not converge, we simplified the random effects structure as suggested by Barr et al. (2013).

Response times. RTs were time-locked to the onset of the prompt screen. Analyses were carried out on log-transformed response times using linear mixed models.

Index of Cognitive Activity. To calculate the ICA we followed the procedure described for Experiment 2 (see Data Analysis in Section 4.1.1).

Fixations. Eye-tracking data were pre-processed as follows. First, because the objects that were used to construct the visual displays differed in size (cf. rucksack vs. mitt in Fig. 4.14), areas-of-interest were calculated per object as the surface that the object covered on the screen in pixels plus 30 pixels around it. Next, fixations shorter than 80 ms 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. Finally, trials with recording problems (e.g., miscalibrations, track loss, etc.) were excluded from the analysis.

For the analysis in the adjective region, in order to account for the difference in the duration of colour and pattern adjectives, we considered a region from 200 ms before adjective offset until 200 ms after noun onset, since it is known that it takes around 200 ms to plan and execute a saccade (Matin et al., 1993). As discussed above, the Specificity manipulation is not relevant on the adjective, as it is determined by the noun. We therefore collapsed across the MS and OS conditions, and coded looks to singleton vs. contrast objects to estimate whether participants assigned a contrastive reading to the prenominal adjective.

For the analysis in the noun region, we were interested in the influence of Specificity and Entropy Reduction on fixating the target referent, and not in possible early effects (anticipatory eye movements are analysed in the adjective region). We therefore considered fixations that started between 300 and 800 ms after noun onset.

In both regions, we considered mean log-gaze probability ratios (see Knoeferle and Kreysa, 2012) of participants' fixations to (a) the *singleton* over the *contrast* object in the adjective region and (b) the *target* over the *competitor* object in the noun region.[2] A positive ratio for (a) would indicate that the singleton object was more likely to be fixated over the contrast, and a positive ratio for (b) that the target object was more likely to be fixated over the competitor. Negative values should be interpreted in the opposite way (i.e., as more looks to the contrast object in the adjective region, and as more looks to the competitor object in the noun region). A score of zero would indicate no differences in the probability with which each object was fixated.

[2]Recall that the *target* was part of the contrast pair in the MS conditions, while was singleton in the OS conditions (cf. the blue ball in Fig. 4.14). The *competitor* was the other referent that matched the description up to the adjective, and was singleton in the MS conditions, but was part of the contrast pair in the OS conditions (cf. the blue mitt in Fig. 4.14).

Because the log ratios are based on aggregation, it is not possible to include crossed random effects of participants and items in the same model. We, therefore, fitted separate linear mixed effects models over participants and over items.

4.2.2 Results

Response accuracy was high in all conditions: Participants gave the correct response at a rate of 95.3% in the MS-HR condition, 95.4% in the MS-LR condition, 96.8% in the OS-HR condition, and 96.5% in the OS-LR condition. We analysed response accuracy using generalised linear mixed models (lme4 package; Bates et al., 2015 in R (version 3.5.1; R Core Team, 2018) with a binomial function, including Entropy Reduction, Specificity and Feature as fixed factors. Because the maximal model including crossed random effects for both participants and items did not converge, the random effects structure was simplified as suggested by Barr et al. (2013). Results showed a main effect of Entropy Reduction ($\beta = 0.141$, $SE = 0.037$, $z = 3.781$, $p < .001$) such that response accuracy was higher in the HR compared to the LR conditions, and an interaction between Entropy Reduction and Specificity ($\beta = -0.171$, $SE = 0.085$, $z = -2.014$, $p = .044$) indicating that participants' response accuracy was particularly improved in the OS-HR condition. All analyses presented below included only trials with correct responses.

Response times. All of the factors included in the model significantly influenced RTs. Participants were faster to respond in the HR conditions (611 ms, $SD = 374$) compared to the LR conditions (659 ms, $SD = 397$; $\beta = -0.0796$, $SE = 0.0155$, $t = -5.14$, $p < .001$), and faster in the OS conditions (614 ms, $SD = 372$) compared to the MS conditions (656 ms, $SD = 398$; $\beta = 0.058$, $SE = 0.016$, $t = 3.755$, $p < .001$). Faster responses were further observed when the mentioned feature was colour (570 ms, $SD = 323$) compared to pattern (703 ms, $SD = 432$; $\beta = -0.192$, $SE = 0.027$, $t = -7.217$, $p < .001$). In addition, the three-way interaction between Entropy Reduction, Specificity and Feature significantly influenced RTs ($\beta = 0.135$, $SE = 0.062$, $t = 2.181$, $p < .05$). We followed up this interaction by fitting separate models for colour and pattern items, and we observed similar results. In colour items, RTs were faster in the HR conditions (545 ms, $SD = 306$) compared to LR conditions (594 ms, $SD = 338$; $\beta = -0.086$, $SE = 0.020$, $t = -4.235$, $p < .001$), and faster in the OS conditions (555 ms, $SD = 323$) compared to MS conditions (584 ms, $SD = 323$; $\beta = 0.053$, $SE = 0.020$, $t = 2.651$, p

< .01). Similarly in pattern items, RTs were faster in the HR conditions (679 ms, *SD* = 423) compared to the LR conditions (726 ms, *SD* = 439; β = -0.073, *SE* = 0.023, *t* = -3.147, $p < .01$), and faster in the OS conditions (676 ms, *SD* = 409) compared to the MS conditions (729 ms, *SD* = 452; β = 0.064, *SE* = 0.023, *t* = 2.773, $p < .01$). The Entropy Reduction x Specificity interaction was marginally significant (β = -0.078, *SE* = 0.046, *t* = -1.688, *p* = .092), such that RTs were slower in the MS-LR condition.

Index of Cognitive Activity. In the adjective time-window (see Fig. 4.16), the Entropy Reduction manipulation was found to significantly influence cognitive effort, with higher ICA values in the HR vs. the LR conditions (β = -0.026, *SE* = 0.013, *z* = -2.068, *p* = .039). The effect of Feature and the interaction between the two factors did not reach significance ($p > .05$).

In the noun region, all of the factors significantly affected participants' cognitive workload (Fig. 4.17). Specifically, we again observed a main effect of Entropy Reduction, this time with higher ICA values in the LR compared to the HR conditions (β = -0.073, *SE* = 0.023, *z* = -3.160, $p < .01$). Furthermore, Specificity and Feature were also found to be significant predictors of cognitive load, with higher ICA values for

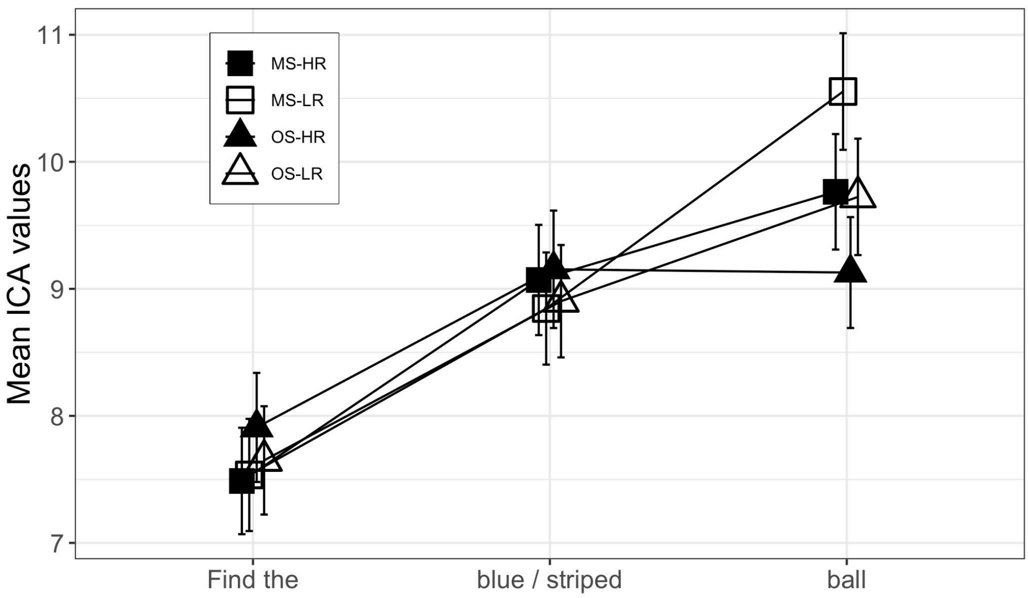


Figure 4.16 Experiment 3. Mean ICA values per condition and region. Error bars represent 95% CIs.

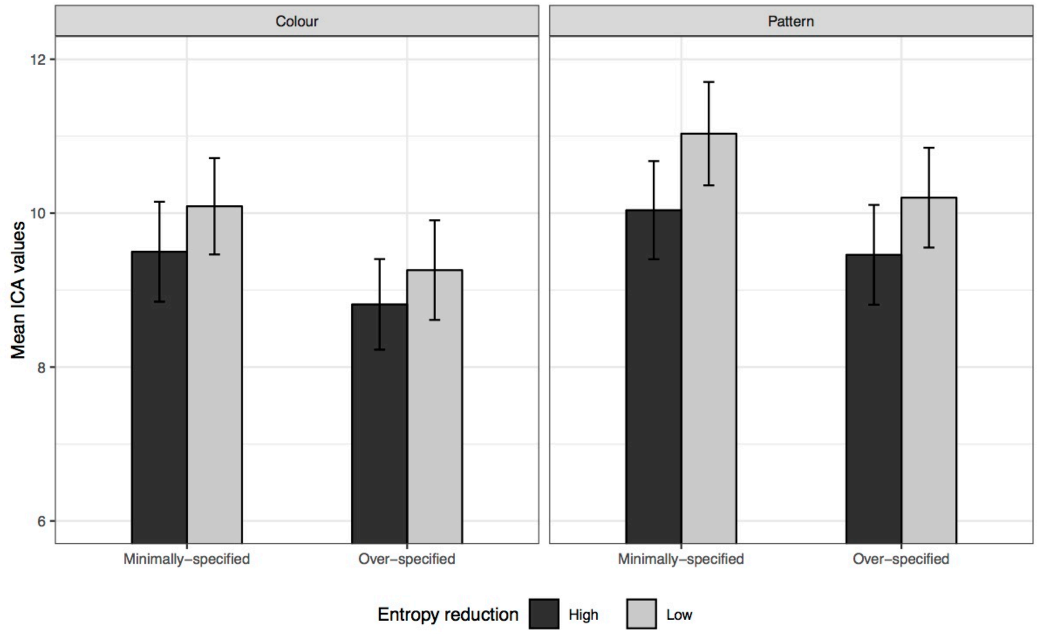


Figure 4.17 Experiment 3. Mean ICA values per condition in the noun region. Error bars represent 95% CIs.

the MS compared to the OS conditions (β = 0.079, SE = 0.026, z = 3.069, $p < .01$), and for Colour compared to Pattern items (β = -0.076, SE = 0.022, z = -3.372, $p < .001$). None of the interactions reached significance (p > .05).

Log-gaze probabilities. As already explained, in the adjective time-window we collapsed across Specificity, and included only Entropy Reduction and Feature as fixed factors in the models. We computed log-gaze probability ratios comparing fixations to the *singleton* and *contrast* objects. Table 4.2 presents the results of this analysis. As indicated by the significant intercept (both by participants and by items), upon hearing the adjective participants were more likely to fixate the contrast object over the singleton object (see negative coefficient). This viewing pattern seemed to be modulated by an interaction between the rate of entropy reduction and the mentioned feature, which we followed up with separate analyses for Colour and Pattern items. In Colour items (see Fig 4.18), none of the comparisons reached significance; there was only a marginal effect on the intercept in the by-participants analysis. In Pattern items (see Fig 4.19), the contrast object was more likely to

Table 4.2 Experiment 3. Log-gaze probability results – Adjective region

	by participants				by items			
	β	SE	t	p	β	SE	t	p
All items								
Intercept	-0.243	0.081	-3.000	.006**	-0.168	0.077	-2.176	.031*
Reduction	-0.111	0.119	-0.926	.364	-0.093	0.139	-0.664	.508
Feature	0.197	0.112	1.756	.091.	0.228	< 0.154	1.481	.141
Reduction:Feature	0.583	0.301	1.938	.064.	0.483	0.279	1.732	.086.
Colour items								
Intercept	-0.144	0.073	-1.985	.059.	-0.053	0.111	-0.482	.632
Reduction	0.183	0.178	1.030	.313	0.149	0.210	0.710	.481
Pattern items								
Intercept	-0.342	0.112	-2.882	.008**	-0.283	0.107	-2.626	.011*
Reduction	-0.402	0.205	-1.960	.062.	-0.335	0.182	-1.838	.071.

Coefficients, SE, *t*- and *p*-values for log-gaze ratios of fixations to singleton vs. contrast objects. *** p<.001, ** p<.01, * p<.05, . p<.1

be fixated over the singleton, and this effect was stronger in the HR vs. the LR conditions.

In the noun region, we considered fixations to the target vs. the competitor object, and included Specificity as a predictor in the models. The results of these analyses are presented in Table 4.3. Even though the analysis by participants resulted only in a marginally significant three-way Specificity x Reduction x Feature interaction, and no other comparison reached significance, several significant effects were found in the by-items analysis. First, there was an effect of Specificity with more looks to the target over the competitor object in the OS vs. the MS conditions. We also found an effect of Reduction such that the target was more likely to be fixated than the competitor in the HR vs. the LR conditions, and an effect of Feature with more fixations to the target in Colour vs. Pattern items. Additionally, there was a significant Specificity x Feature interaction with more fixations to the target object in the OS conditions for Colour items. We followed up the interactions by performing separate analyses for Colour and Pattern items. In Colour items (see Fig. 4.20), the by-participant analysis resulted only in a marginally significant effect of Specificity, with more looks to the target object in the OS conditions. The by-items analysis revealed a significant effect of Specificity in the same direction, and a significant

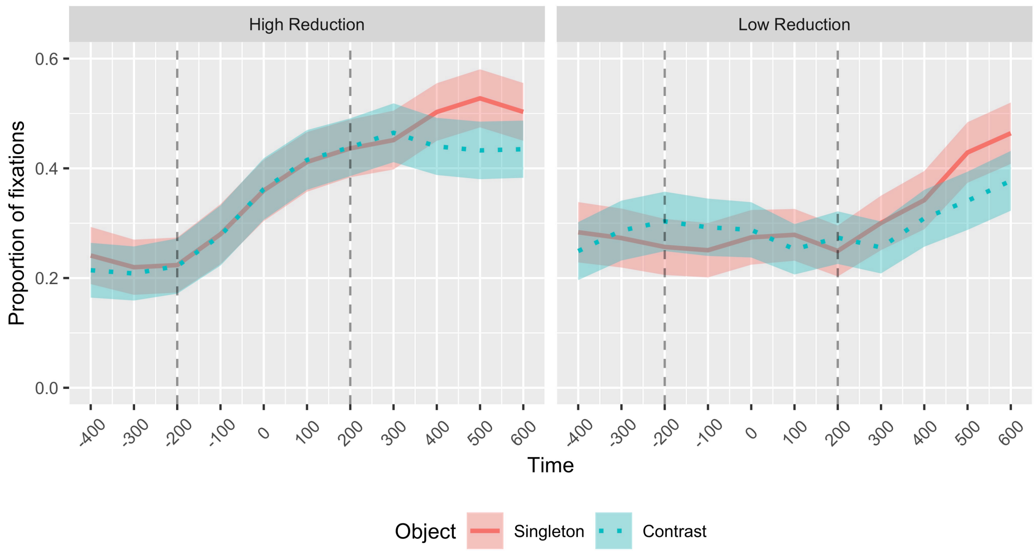


Figure 4.18 Experiment 3. Proportions of fixations to singleton vs. contrast objects in Colour items, from –200 ms until 200 ms around noun onset (see dashed lines). The shaded bands represent 95% CIs.

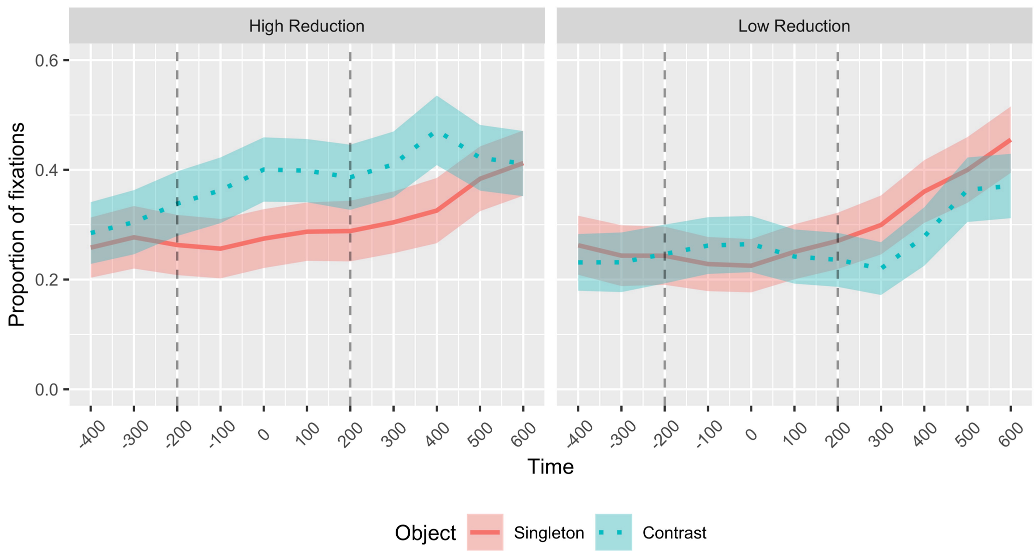


Figure 4.19 Experiment 3. Proportions of fixations to singleton vs. contrast objects in Pattern items, from –200 ms until 200 ms around noun onset (see dashed lines). The shaded bands represent 95% CIs.

Table 4.3 Experiment 3. Log-gaze probability results – Noun region

	by participants				by items			
	β	SE	t	p	β	SE	t	p
All items								
Intercept	0.773	0.056	13.825	< .001***	0.745	0.037	20.358	< .001***
Specificity	-0.097	0.066	-1.463	.157	-0.138	0.054	-2.550	.012*
Reduction	0.070	0.056	1.345	.192	0.129	0.065	1.999	.048*
Feature	0.052	0.062	0.830	.415	0.159	0.073	2.173	.032*
Spec:Red	0.192	0.120	1.606	.122	0.153	0.123	1.238	.218
Spec:Feat	-0.165	0.163	-1.016	.320	-0.229	0.108	-2.120	.036*
Red:Feat	0.023	0.101	0.226	.823	0.128	0.130	0.989	.325
Spec:Red:Feat	-0.239	0.129	-1.849	.065.	-0.324	0.247	-1.315	.191
Colour items								
Intercept	0.798	0.071	11.283	<.001***	0.825	0.056	14.765	<.001***
Specificity	-0.179	0.099	-1.815	.083.	-0.252	0.084	-2.990	.004**
Reduction	0.082	0.072	1.138	.267	0.193	0.087	2.213	.031*
Spec:Red	0.073	0.162	0.447	.659	-0.009	0.177	-0.055	.957
Pattern items								
Intercept	0.747	0.056	13.243	<.001***	0.666	0.047	14.067	<.001***
Specificity	-0.014	0.110	-0.128	.899	-0.023	0.067	-0.346	.730
Reduction	0.058	0.073	0.802	.430	0.065	0.096	0.687	.495
Spec:Red	0.311	0.141	2.213	.037*	0.314	0.171	1.833	.072.

Coefficients, SE, *t*- and *p*-values for log-gaze ratios of fixations to singleton vs. contrast objects. *** p<.001, ** p<.01, * p<.05, . p<.1

effect of Entropy Reduction with more looks to the target over the competitor in the HR vs. LR conditions. In Pattern items (see Fig. 4.21), the Specificity x Reduction interaction was significant in the analysis by-participants and marginally significant in the analysis by-items. This interaction seemed to be driven by a smaller log-ratio in the MS-LR condition compared to the other three conditions (see Table 4.4).

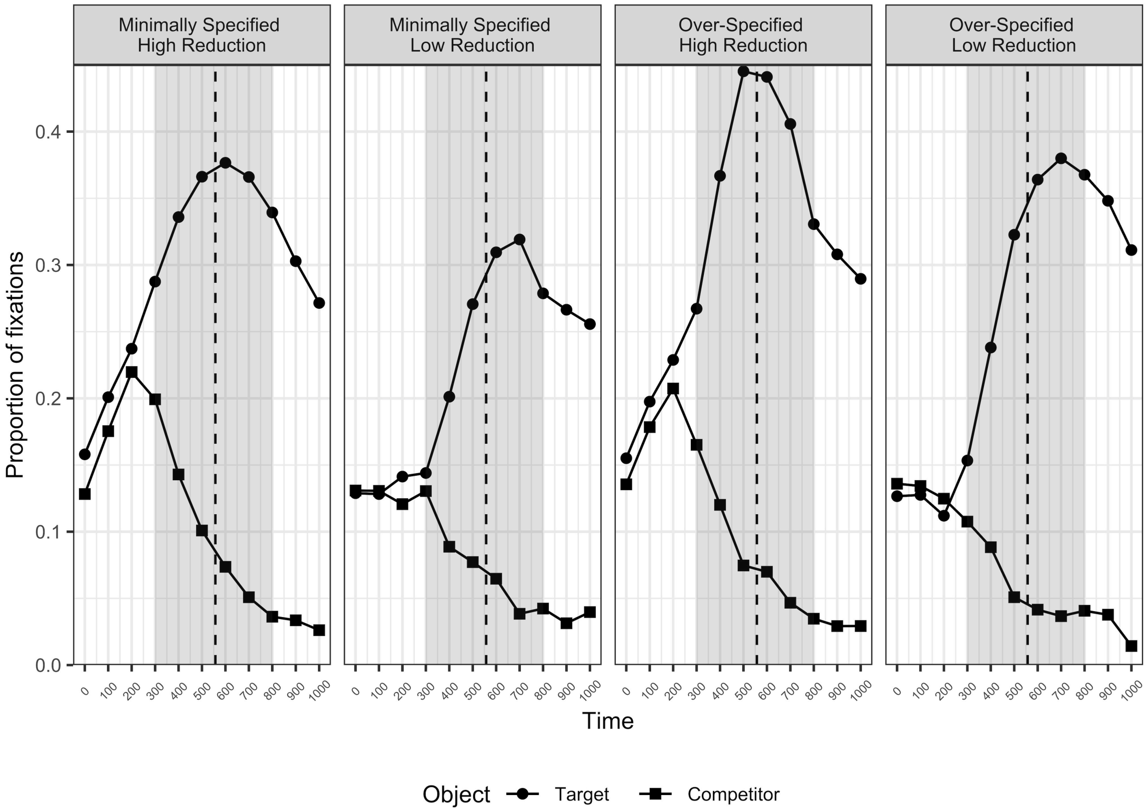


Figure 4.20 Experiment 3. Proportions of fixations to the target vs. competitor objects in each condition for Colour items. The shaded area represents the 300-800 ms post noun onset analysis time-window.

Table 4.4 Experiment 3. Mean log-gaze probability ratios for fixations to the target object over fixations to the competitor object in the noun region of pattern items

	Mean log ratios by participants	Mean log ratios by items
MS-HR	0.847 (0.729)	0.765 (1.011)
MS-LR	0.632 (0.821)	0.548 (1.038)
OS-HR	0.705 (0.814)	0.636 (1.049)
OS-LR	0.802 (0.800)	0.726 (1.015)

Note: SD in parentheses

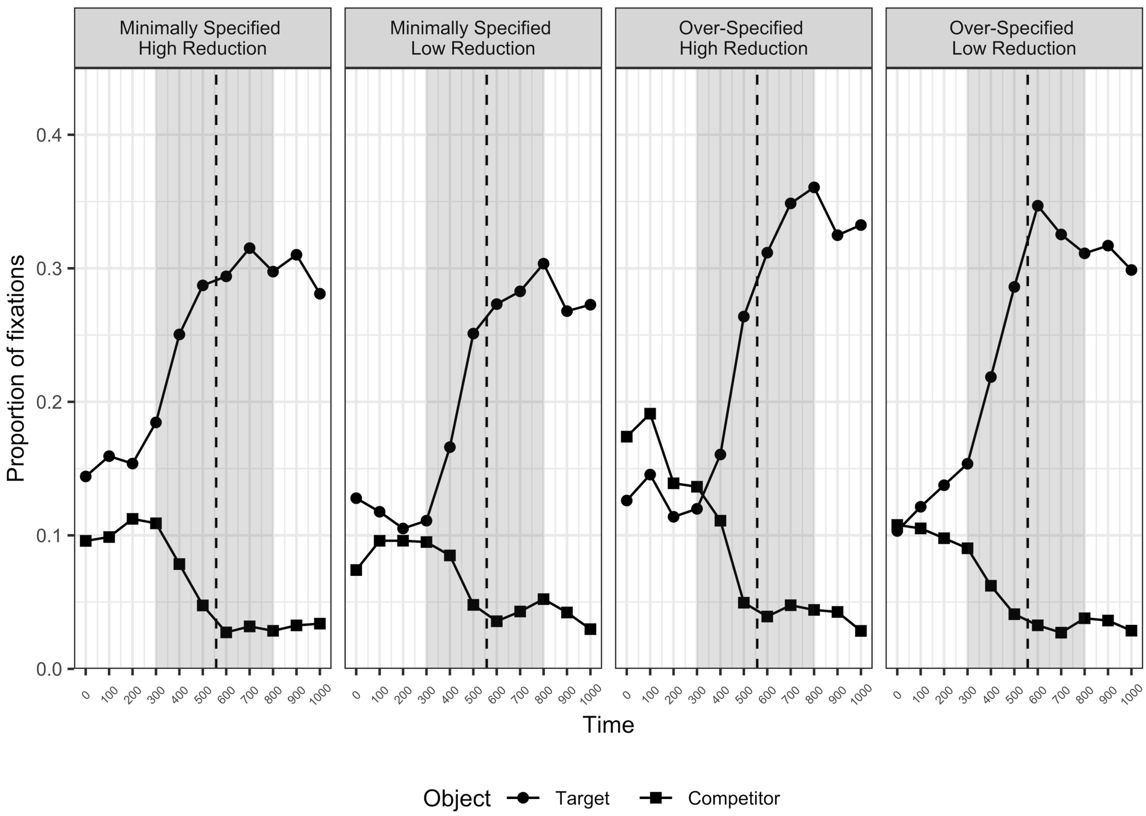


Figure 4.21 Experiment 3. Proportions of fixations to the target vs. competitor objects in each condition for pattern items. The shaded area represents the 300-800 ms post noun onset analysis time-window.

4.2.3 Discussion

Experiment 3 aimed at replicating the findings from Experiment 2, measuring participants' eye movements and ICA. More specifically, we asked (a) to what extent the rate of Entropy Reduction in the expression influences processing and (b) whether and how, above and beyond any effects of Entropy Reduction, the Specificity of the referring expression may affect comprehension in visual contexts that support both a contrastive (MS conditions) and a non-contrastive (OS conditions) interpretation of the adjective. Moreover, we aimed at gaining a better estimate of the ICA measure, by recording the ICA while participants could freely scan the visual scene.

Regarding Specificity, this manipulation became relevant in the noun region, when the type of the target referent was revealed. In this region, participants' ICA values were lower in the OS compared to the MS conditions, indicating that over-specifications in fact facilitated comprehension, contrary to the findings from Experiment 2. This evidence was further supported by the log-gaze probability ratios. Unsurprisingly, participants looked more towards the target object vs. the competitor object (i.e., the other referent that matched the description up to the adjective; cf. the blue mitt in Fig. 4.14) after hearing the noun. However, this effect was modulated by Specificity, such that looks to the target over the competitor were more likely when the noun followed a redundant adjective (OS conditions) compared to when it followed a necessary adjective (MS conditions) (recall that in the MS but not in the OS conditions a referent of the same type as the target is available on the opposite side of the visual scene, possibly interfering with target identification; cf. the green ball in 4.14 a and b). Furthermore, participants' RTs were found to be faster in the OS vs. the MS conditions. Crucially, these results were similar for both Colour and Pattern items.

In the adjective region, we expected anticipatory looks to the singleton vs. the contrast objects to reflect participants' interpretation of the adjective. Whereas we found some evidence that the contrast object was fixated more than the singleton (supporting the Gricean account), this only occurred with pattern adjectives. Pattern is, however, generally more difficult to discern than colour, and even more so in our stimuli set, where, even though colours were matched across object types, patterns were not (e.g., the dots on a dotted object could be larger compared to another dotted object on the same scene; see object pictures in Appendix A). It is, therefore, possible that this effect is associated to the relative ease of discriminability between colour and pattern items. Additionally, the length of pattern adjectives, which in this experiment were on average 200 ms longer than colour adjectives, may have contributed to the differing results between Colour and Pattern items. Thus, with pattern adjectives participants may have had more time to consider which object could be the target referent and to employ Gricean reasoning. Nevertheless, participants' gaze pattern in the adjective region of colour items, as well as the facilitation (lower ICA) found for over-specifications with both Colour and Pattern items in the noun time-window, contradict the Gricean account, and support the view that over-specification facilitates comprehension. A particularly interesting finding in support of this claim comes from the comparison of the ICA measurements in the OS-HR trials and their MS-filler counterparts. Recall that a large part of the filler trials were essentially the minimally-specified versions of the OS experimental trials. For example, visual scenes such as in Figure 4.14c would be paired with the phrase 'Find the ball', i.e., without an adjective. In a post-hoc analysis, we included the MS-filler, and compared the ICA values in the noun region between the OS-HR condition and the MS-filler. The findings indicated that the ICA values were significantly lower in the OS-HR condition compared to the MS-filler (β = -0.078, SE = 0.039, z = -1.989, p = .046). This result suggests that listeners' cognitive effort on the final noun was facilitated when this word followed a redundant adjective (regardless of the kind) compared to when it was a bare noun (which would be the case if the instruction was minimally-specified in the *same* visual scene).

As for Entropy Reduction, our findings generally support the Entropy Reduction Hypothesis (Hale, 2003, 2006), and show that the reduction of uncertainty is a predictor of comprehension difficulty in visually-situated communication. In contrast to Ankener et al. (2018), we found effects of entropy reduction at each reduction point in the ICA measurement. A high reduction of entropy on the adjective resulted in higher cognitive effort in that region, but facilitated processing on the following noun; residual entropy on the noun – and the cognitive effort associated with the reduction of this entropy – was lower in HR than in LR trials. The facilitation for the HR vs. LR conditions observed on the noun was further supported by the increased likelihood to fixate the target over the competitor object during the noun in the HR compared to the LR conditions, as well as by faster RTs in HR compared to LR trials.

In sum, despite a preference for interpreting pattern adjectives contrastively (anticipatory eye movements), Experiment 3 found evidence that referential redundancy generally benefits processing (ICA, log-gaze probabilities, RTs). Furthermore, a high reduction of entropy on the adjective was found to increase effort in that region, but it facilitated processing on the subsequent noun (ICA). These effects were similar across Colour and Pattern items, but, overall, redundant colour adjectives resulted in greater facilitation.

4.3 Summary

Experiments 2 and 3 addressed RQ-2, asking whether the rate of entropy reduction across a referring expression influences processing, and whether any effect of over-specification would be additive to that of entropy reduction. We, furthermore, re-visited RQ-1, asking how redundancy affects listeners' on-line comprehension of referring expressions in visual contexts that support both a contrastive and a non-contrastive reading of a (redundant) adjective – i.e., when the adjective can be interpreted as referring to either a contrast or a singleton object. In both experiments we, therefore, manipulated Specificity and Entropy Reduction as orthogonal factors, and recorded listeners' ERPs (Experiment 2), and eye movements (Experiment 3). For comparability between the results of the two experiments, we additionally measured participants' ICA in both of them. Our findings offer important insights, that are interesting both from a methodological and a theoretical point of view.

First, the results of the two experiments combined contribute to an interesting conclusion regarding the interpretation of the ICA. Although in both experiments we used the same set of stimuli and followed the same procedure, only in Experiment 3 was the experimental manipulation found to modulate the ICA. The only difference between the two experiments was that in Experiment 3 participants could freely scan the visual scene when processing the referring expression, while in Experiment 2 they had to fixate a cross in the middle of the screen (due to the concurrent registration of the EEG signal). We, therefore, believe that the ICA is sensitive to cognitive workload induced by *visually-grounded* language processing. In other words, the cognitive effort that the ICA measures is related to the allocation of attention across the visual scene. Therefore, when participants' gaze needs to be fixed for some period of time (as in Experiment 2) – inhibiting shifts in overt attention during this period – the ICA appears to be neutralised.

Regarding the comprehension of over-specifications, our results generally indicate that redundant colour and redundant pattern adjectives are processed differently. More specifically, as shown with anticipatory eye movements, listeners expected pattern adjectives to refer to objects in contrast pairs (MS conditions), rather than to singleton objects (OS conditions). In other words, pattern adjectives are likely to be interpreted contrastively. This finding was further supported by our ERP results, showing that processing of the noun was hindered after a redundant pattern adjective (N400 effect for the OS vs. MS conditions in Pattern items). These results support the Gricean account, according to which listeners expect speakers to deliver the precise amount of information that is required to identify a referent, and when this is not the case listeners' referential processing is hindered. Nevertheless, participants' cognitive effort on the noun, as measured by the ICA, was lower after a redundant than after a necessary pattern adjective (i.e., lower in the OS compared to the MS conditions). As argued above, we believe that the ICA taps into visually-grounded processing effort. We, therefore, take the lower ICA values observed for the OS conditions to suggest that, regardless of how redundant pattern adjectives are incrementally interpreted, they help guide listeners' visual search for the target referent narrowing down the set of potential referents, thereby decreasing cognitive workload on the noun. This interpretation is further supported by (a) the increased P600 observed for the OS vs. the MS conditions on the noun, suggesting that it was easier for participants to perform the task (make a button press to indicate which side of the screen the target referent was on) in the OS conditions, and (b) the behavioural data, showing that participants were faster to identify the position of the target referent on the screen, when the instructions were over-specified compared to minimally-specified.

Regarding redundant colour adjectives, we found no evidence that they were interpreted contrastively (no anticipatory eye movements), or that they hindered processing of the upcoming noun (no N400). Redundant colour adjectives did, nevertheless, decrease cognitive load on the noun (lower ICA values), facilitating task performance, especially in the OS-HR condition (P600 effect), and speeding up response times. Taken together, the results regarding Specificity indicate that, even though incrementally pattern adjectives were assigned a contrastive interpretation and colour adjectives were interpreted non-contrastively, redundant adjectives in any case facilitated participants' visual search for the target referent and enhanced their performance in the task.

Both experiments converge on the finding that Entropy Reduction is a predictor of processing effort in visually-situated comprehension, regardless of the kind of adjective used. A higher reduction of entropy on the adjective facilitated processing on the following noun (reduced N400). Additionally, Entropy Reduction affected participants' cognitive load induced by the visual search for the target referent: ICA values for the HR vs. LR conditions were higher on the adjective, where referential entropy is reduced more radically, but lower on the noun, where residual entropy is lower in the HR compared to the LR conditions. These findings are further supported by our RTs results, which show that participants were faster to identify the target referent's location when referential entropy on the final noun was lower (in the HR vs. LR conditions). These results are in line with the Entropy Reduction Hypothesis (Hale, 2003, 2006), and the bounded-rational account of over-specification, according to which redundant adjectives may even benefit visually-situated comprehension, when they help restrict the set of potential referents and ease listeners' identification of the target.

Summarising, results from Experiments 2 and 3 show that, although redundant pattern adjectives may hinder processing, redundant colour adjectives do not. In any case, both kinds of redundant adjectives facilitate visual search for the target, as they help restrict the set of potential referents and reduce uncertainty about the target referent. The rate at which this referential entropy is reduced across the utterance was found to further influence comprehension, with a high reduction of entropy on the adjective resulting in difficulty in that region, but in a facilitation on the following noun. We now turn to the production of referring expressions, and the question whether speakers are sensitive to these processing concerns and take comprehension effort into account when planning their utterances.

Chapter 5

The influence of referential entropy on the production of over-specifications

The experiments described thus far aimed at understanding how over-specified referring expressions are processed by listeners, and at examining whether the rate of referential entropy reduction in these expressions affects comprehension. Taken together, our results offer a few important insights. First, they show that the comprehension of over-specifications largely depends on the kind of redundant adjective used in the expression: While redundant *pattern* adjectives were interpreted contrastively thus resulting in difficulty with processing the subsequent noun, no such effects were observed for redundant *colour* adjectives, whose processing did not differ from that of necessary adjectives (cf. anticipatory eye movements). Both kinds of redundant adjectives, however, facilitated listeners' visual search for the target referent and assisted their identification of the intended referent. Secondly, the rate at which a prenominal adjective reduced referential entropy significantly influenced processing, as predicted by the Entropy Reduction Hypothesis (Hale, 2003, 2006): A high reduction of entropy on the adjective resulted in increased effort associated with excluding more objects from consideration in that region, but in facilitation on the subsequent noun. Therefore, our findings so far generally support the bounded-rational account, according to which any word in an utterance, even if it is redundant, will benefit comprehension, as long as it helps reduce referential entropy more efficiently (i.e., quickly, easily and accurately).

Given these results, Experiments 4 and 5 presented in this chapter aimed at identifying which factors motivate the *production* of over-specifications, and whether these factors are primarily associated with *egocentric* or *addressee-oriented* (whether Gricean or bounded-rational) concerns.[1] The egocentric view holds that production preferences are tuned to minimise speakers' effort, regardless of the addressees' needs. Therefore, if over-specifications are the result of egocentric production processes, speakers' choices should not be affected by the experimental manipulation (i.e., egocentric speakers' OS rate should not vary across conditions). By contrast, according to the addressee-oriented view, speakers should prefer structures that ease comprehension for their listeners. Both the Gricean and the bounded-rational approaches are in agreement with this view, but make different predictions with regard to the rate of speakers' over-specifications. Based on the Gricean view, speakers are expected to produce only minimal descriptions, as redundancy may lead their listeners to derive unintended implicatures and compute a meaning, which they will later need to revise. On the other hand, the bounded-rational approach predicts that speakers may use redundant adjectives when they can ease their listeners' task, e.g., the search for the target object. Such production preferences would be in line with the findings from Experiments 2 and 3 that listeners favour utterances that reduce referential entropy more uniformly. Finally, based on the results from Experiments 2 and 3, as well as on previous research (cf. Rubio-Fernández, 2016, 2019; Sedivy, 2003), colour over-specifications should be expected to be more frequent than pattern over-specifications.

In two referential communication experiments, we investigated whether and how the distributional properties of the visual scene influence the production of referential over-specification by carefully manipulating the potential of a word to reduce entropy (uncertainty regarding the target referent; cf. Hale, 2003, 2006). Identifying which property is more entropy reducing, in order to include it in a description, is arguably more demanding for the speakers than just relying on simple heuristics, such as mentioning the most salient feature. Our general hypothesis, therefore, is that over-specifications that include the most informative property —- in terms of uncertainty reduction -— aim at making visual search more effective for addressees and thus facilitate referential communication.

In both experiments, pairs of participants were seated in front of two monitors and were presented with arrays of objects. While the objects were the same for

[1] Parts of this chapter were published in a journal article (see Tourtouri et al., 2019).

both participants, their arrangement in the array differed in half of the displays. Participants had to collaborate in order to identify whether an object marked as the target on the Speaker's display was in the same position on the Listener's display as well. Objects differed in colour and pattern, but on critical trials they were not both necessary to specify the target. In Experiment 4, the target referent was singleton and therefore did not require modification in order to be specified, while in Experiment 5 it was part of a contrast pair and required the mention of either colour or pattern. In both experiments, we manipulated which of the target referent's properties restricted referential search space at a higher degree, thereby reducing uncertainty about the target referent. In other words, we manipulated whether the target colour or pattern could reduce referential entropy more, or whether they were both equally entropy-reducing. The two experiments were run simultaneously with critical trials from one serving as fillers for the other. We measured the proportion of over-specifications produced per condition.

5.1 Experiment 4

In Experiment 4, we examined speakers' use of redundant adjectives in visual contexts where the target referent was singleton, and therefore no adjective was necessary to specify it (cf. the ball in Fig. 5.1). The experimental manipulation concerned whether one of the target features (colour or pattern) was *more entropy-reducing* than the other. That is, in critical displays it could be that colour reduced entropy at a higher rate compared to pattern (see Fig. 5.1a and d), pattern reduced entropy at a higher rate compared to colour (see Fig. 5.1b and e), or that both features reduced entropy at a similar rate (see Fig. 5.1c and f). As the results from the previous comprehension experiments suggest, listeners may expect adjectives – especially pattern adjectives – to be used contrastively. Given that, speakers may minimise their use of redundant modifiers, in case these modifiers also select an object from a contrast pair, in order not to misguide their listeners' interpretation of the unfolding utterance. Since the aim of this experiment was to examine what motivates speakers' use of *redundant* adjectives, in experimental displays the competitor object was the only other singleton that matched with the target referent in both features (cf. the mitt in Fig. 5.1). This was done so that the use of a redundant adjective would not allow listeners to select the target immediately after hearing the adjective, and that speakers would therefore not refrain from over-specifying.

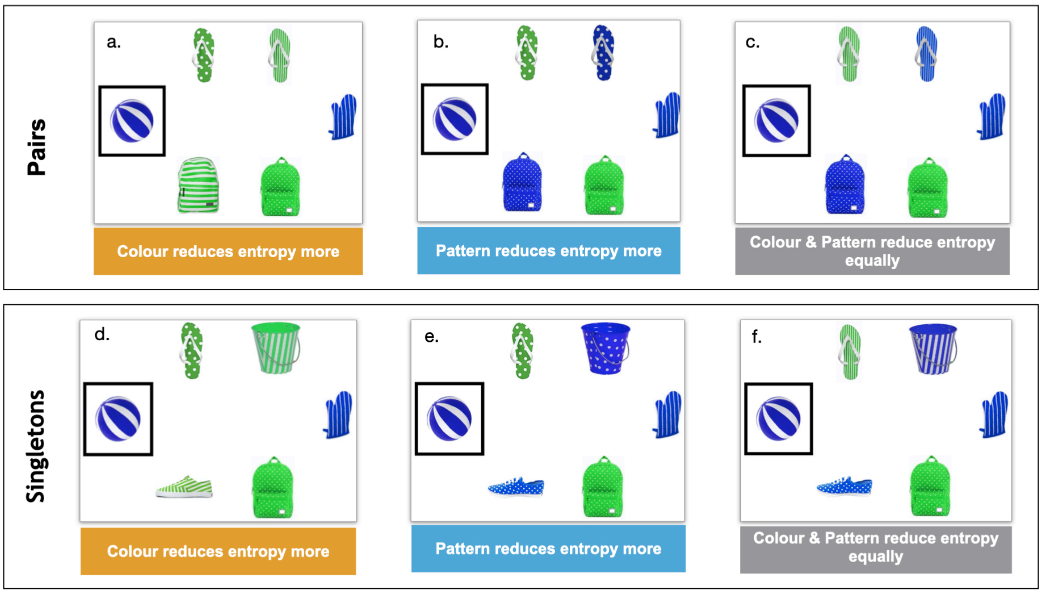


Figure 5.1 Experiment 4. Sample visual scenes (conditions) for an experimental item.

Summarising the hypotheses above, if speakers' production choices are *egocentric*, OS rates should be similar in all conditions. By contrast, if speakers' choices are motivated by audience-design concerns, results may be in agreement either with the Gricean or with the bounded-rational approach. According to the *Gricean* approach, speakers are expected to use only unmodified utterances to refer to the target object, as listeners should expect adjectives to mark a distinction between similar objects. That is, speakers should use the unmodified expression 'the ball' to refer to the intended referent in all panels of Figure 5.1. On the other hand, according to the *bounded-rational* approach speakers may over-specify when the redundant adjective contributes to a smooth entropy-reduction profile. For example, in Figure 5.1a, referential entropy at the beginning of the utterance (before any critical information is produced) is 2.58 bits. An unmodified noun would reduce entropy to 0 bits, which is a reduction of entropy by ΔH_{ball} = 2.58 bits in just one step. If on the other hand, a redundant colour adjective (e.g., 'blue') is used, entropy reduction across the utterance would be more uniform, since the adjective acts as an intermediate reduction step. That is, referential entropy on 'blue' would be reduced by ΔH_{blue} = 1.58 bits, and by ΔH_{ball} = 1 bit on the final word. The use of a redundant *pattern* adjective (i.e., 'striped') in Figure 5.1b reduces entropy in a similar way. By contrast, in Figure 5.1c none of the adjectives can reduce entropy as uniformly, as they both bring about a reduction of $\Delta H_{blue/striped}$ = 0.58 bits, leaving a larger amount of entropy to be reduced on the final noun, i.e., ΔH_{ball} = 2 bits. Therefore, according to the bounded-rational approach we should expect a higher rate of over-specifications for colour (Colour-OS) when colour reduces entropy more than pattern, a higher rate of over-specifications for pattern (Pattern-OS) when pattern reduces entropy more than colour, and a lower OS rate for any of the two adjectives when they both reduce entropy equally.

5.1.1 Method

Participants. Forty-nine pairs of native German speakers, who had no problems with colour perception and had not taken part in any of our previous experiments, participated in this study. They were randomly assigned to the roles of Speaker (36 female, mean age = 23.2) and Listener (33 female, mean age = 24.3). Participants gave informed consent, and were monetarily compensated for their participation.[2] One pair of German-French bilinguals was not included in the analysis due to French language dominance.

Materials. Visual stimuli were created using eighteen of the object pictures from the previous experiments (see Appendix A). One experimental item comprised six versions of one display (see Fig. 5.1), differing in which of the target features (colour or pattern), if any, had a higher entropy-reduction potential. When the target colour selected fewer objects than the target pattern (see Fig. 5.1a and d), colour was more entropy-reducing (Colour-Reducing conditions). When the target pattern selected fewer objects than the target colour (see Fig. 5.1b and c), pattern was more entropy-reducing (Pattern-Reducing conditions). Finally, when both target colour and pattern selected the same number of objects (see Fig. 5.1c and f), entropy reduction was equal for both features (Equally-Reducing conditions). As in the previous experiments, critical displays always contained six objects. The target referent (see the ball in Fig. 5.1) as well as the competitor object (i.e., the referent competing with the target for both features in all conditions; cf. the mitt in Fig. 5.1) were always singletons, and shared both colour and pattern features. In order to

[2] In a few cases, two participants could not be scheduled for the same time slot, therefore confederates were used as Listeners. However, the listener's role in this task is minor and should not influence the speakers' results.

increase variability in the visual stimuli, distractor objects were arranged in contrast pairs in half of the critical displays (see the flip-flops and rucksacks in Fig. 5.1a–c), while they were singletons in the other half (see Fig. 5.1d–f).

A total of 216 experimental displays were created. Another set of 72 filler items was also constructed. Filler displays depicted either six or four objects, again differing in colour and pattern. The target referent in the filler items was either part of a contrast set of three same-type objects or it was a singleton. Thus, fillers required either two adjectives for disambiguation of the intended referent or none. Finally, fillers also varied in whether one of the target properties was more entropy-reducing than the other or both properties reduced entropy to an equal extent.

Experimental and filler displays were intermixed with stimuli from Experiment 5. Overall, 576 visual scenes were created, half of which were then flipped on the vertical axis and were used only on the Listeners' display. The Listeners, therefore, saw half of the items in the same configuration as the Speaker and the other half in a mirrored configuration (recall that participants' task was to identify whether the object marked as the target on the Speaker's display was on the same side of the screen for the Listener, too). Stimuli were distributed into six lists following a Latin square design, so that only one version of an item appeared in a list, and so that participants were exposed to only one condition of each item. Lists were pseudo-randomised so that two trials from the same condition never appeared in a row, and at least one filler and/or one trial from the other experiment intervened between two consecutive experimental trials. E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA) was used to implement and run the experiment.

Procedure. Speakers and Listeners were seated on opposite sides of a glass window separating two adjacent rooms. They each had a 1680 x 1050 resolution monitor in front of them, and used a microphone and headphones to communicate via an audio link. Participants saw displays containing the same objects, but their position on the vertical axis was flipped on half of the trials. They were instructed to imagine taking part in a long-distance call, where they needed to establish whether they share the same visual domain with their partner or not. Their task was to identify whether an object that was designated to the speaker was on the same side of the screen for both of the participants. More specifically, after a 2 second preview time a target object was indicated by a black frame (cf. Fig. 5.1) on the Speaker's screen only, and a sound was played signifying that the target had been revealed to the

Speaker. The Speaker then had to ask the Listener which side of the screen the target object was on. For example, in all panels of Figure 5.1 a question containing minimal information would be 'Is the ball on the left?'. The Listener's task was to respond 'Yes' or 'No' by pressing a button on a Cedrus response pad (Cedrus Corporation, San Pedro, California, USA). Listeners were allowed to ask for further information, if necessary. Feedback was given after each trial in the form of a bell sound (for correct responses) or buzzer sound (for incorrect responses) that was audible to both participants. Crucially, in order to encourage participants to collaborate rather than perform two disjointed tasks, they were told that they only had a limited amount of time to complete each trial.

One experimental session proceeded as follows. When participants came in the lab, they were randomly assigned to the roles of Speaker and Listener. The roles were described as 'Information-seeker' for the Speaker and 'Information-giver' for the Listener, so that participants' behaviour would not be confined by the Speaker/Listener distinction. After participants read the instructions corresponding to their role, the Experimenter explained their tasks in order to ensure that the instructions were clear to both of them, and that they understood that their tasks converged on a common goal. They were then presented with a preview of the objects that would appear during the experiment, in displays arranged by object type and showing all colour and pattern combinations. During this phase, Speakers were asked to name out loud the object type on each display. They were next shown to their seats and completed a practice block. The experiment began after it was confirmed that both participants understood the task. The experimenter remained in the same room as the Speaker during the experimental session, in order to make sure that the Speaker did not use truncated sentences (e.g., 'blue ball left'). Participants reported that the presence of the experimenter did not affect their performance. One experimental session lasted approximately 30 minutes.

Data coding. Speakers' utterances were transcribed and annotated by two trained linguists. Data from two speakers, who produced a high rate of under-specifications in the trials from Experiment 5 and the fillers were excluded from further analyses (see Data coding in Section 5.2.1 for more details), as this suggests that they had not fully understood the task. Data from one more participant, whose audio files were corrupted, were also excluded from analyses. Trials with recording problems and trials where the speaker named a wrong object as the target were excluded from analyses (0.68%). We further excluded trials containing self-repairs (i.e., repairs of the adjective or the noun), or revisions of the utterance structure (i.e., providing more information after a listener request or revising the utterance to delete information) were excluded from analyses (1.63%). Trials with over-specified utterances (44.8%) were coded as 1, and trials with minimally-specified utterances (55.2%) were coded as 0. After we identified speakers who consistently over-specified using both adjectives (8 participants; see below), we excluded trials in which the same kinds of adjectives were used in the same order as in the immediately preceding trial (11.77%), in order to account for potential priming effects from the previous trial. In this way we ensured that over-specification on a particular trial would not merely be the result of priming from the immediately preceding trial that required modification (see trials from Experiment 5 and fillers).

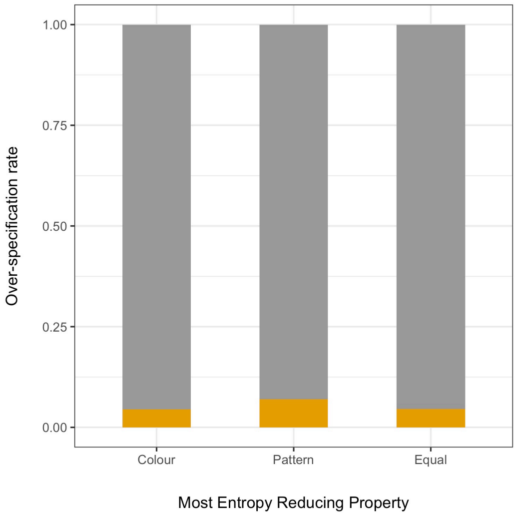
Data analysis. Proportions of over-specifications were analysed using generalised linear mixed models with the lme4 package (Bates et al. 2015) in R (version 3.5.1; R Core Team 2018), including crossed random intercepts for participants and items, and random slopes for the Entropy-Reducing factor. When the maximal models did not converge, the random effects structure was simplified (Barr et al. 2013).

5.1.2 Results

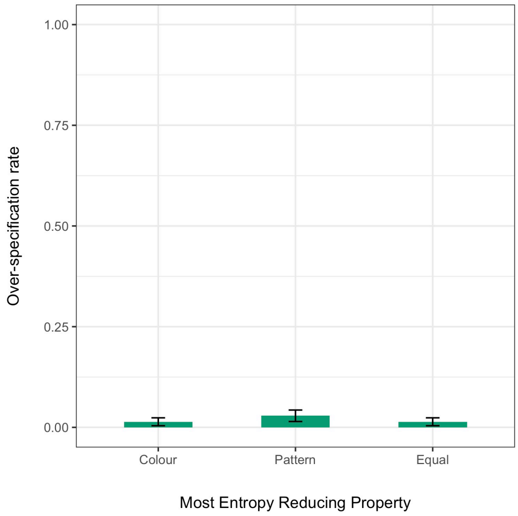
According to the egocentric view, OS rates should be unaffected by the experimental manipulation. Indeed, 8 participants were found to over-specify for both attributes the majority of the time. In line with the Gricean approach, 13 more speakers were found to primarily produce minimal descriptions. We therefore categorised participants into three groups based on their use of redundancy: Group 1 consisted of speakers who used both adjectives redundantly more than 80% of the time (N = 8; Fig. 5.2a). Group 2 consisted of speakers who produced minimal descriptions more than 90% of the time (N = 13; Fig. 5.2b).[3] Finally, Group 3 comprised the remaining participants (N = 24; see Fig. 5.2c). Analyses were performed per group.

In Group 1, analyses were performed on the proportions of over-specifications for both adjectives (Both-OS). No significant differences were found in the proportions

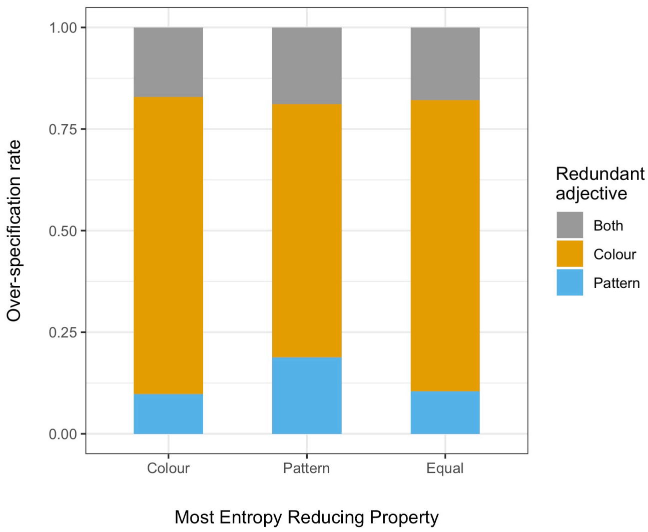
[3] Note that the bars in Figure 5.2b represent the total occurrences of over-specification (shown using a different colour) and are not split for the usage of colour vs. pattern adjectives.



(a) Group 1 (N=8)



(b) Group 2 (N=13)



(c) Group 3 (N=24)

Figure 5.2 Experiment 4. Proportions of over-specifications produced in each group of participants. Error bars in (b) represent SE.

of Both-OS, neither between the Colour-Reducing and Pattern-Reducing conditions (β = 0.509, *SE* = 0.696, *z* = 0.731, *p* = .465) nor between the Colour-Reducing and Equal-Reducing conditions (β = 0.033, *SE* = 0.754, *z* = 0.044, *p* = .965).

Due to the very scarce occurrences of over-specification, none of the models converged in the analyses for Group 2. This fact in combination with the graph in Figure 5.2b make evident that speakers in this group practically never over-specified in any of the conditions.

In Group 3, we performed analyses on the proportions of Colour-OS and Pattern-OS. Regarding Colour-OS, the comparison between the Colour-Reducing and Pattern-Reducing conditions was significant (β = -0.5, *SE* = 0.232, *z* = -2.152, *p* = .031), with a higher rate of Colour-OS in the Colour-Reducing condition. Similarly, the comparison between the Colour-Reducing and Equally-Reducing conditions was also significant (β = -0.465, *SE* = 0.231, *z* = -2.011, *p* = .044), again with a higher proportion of Colour-OS in the Colour-Reducing condition. For Pattern-OS, none of the comparisons reached significance: The proportion of Pattern-OS did not differ between the Pattern-Reducing (reference level) and Colour-Reducing conditions (β = -0.539, *SE* = 0.512, *z* = -1.053, *p* = .292), nor did it differ between the Pattern-Reducing and Equally-Reducing conditions (β = -0.513, *SE* = 0.528, *z* = -0.971, *p* = .331).

5.1.3 Discussion

Experiment 4 examined speakers' tendency to include redundant adjectives in their referring expressions, and whether referential entropy may influence this tendency in visual scenes where the intended referent was a singleton object. Pairs of naïve participants viewed visual displays containing six objects, one of which was designated as the target only to the Speaker. While the target object was always a singleton and, therefore, did not require a modifier in order to be specified, it could be that one of its features (colour or pattern) had a higher entropy reduction potential. The experimental manipulation concerned whether colour or pattern was more entropy-reducing (Colour-Reducing vs. Pattern-Reducing) or whether they both reduced entropy equally (Equally-Reducing). We measured how frequently speakers used a minimally-specified (coded as 0) vs. an over-specified (coded as 1) expression to refer to the target object per condition. The egocentric view predicted that speakers' OS rate should be unaffected by the experimental manipulation and therefore invariable across conditions. According to the Gricean approach, the OS rate should be close to zero, as speakers should prefer minimal descriptions in order not to misguide the listeners' interpretation of the utterance. Lastly, under the bounded-rational approach speakers were expected to use one of the features redundantly, when it was more entropy-reducing than the other.

Three groups of speakers were identified based on their use of over-specifications. Group 1 consisted of participants who used both colour and pattern adjectives redundantly more than 80% of the time. As predicted by the egocentric view, analyses for Group 1 found no differences in the OS rate across conditions: Over-specifications for both features were equally frequent in all conditions. Another two groups of speakers were found to be more listener-oriented. In line with the Gricean approach, speakers in Group 2 almost never over-specified; their OS rate was lower than 10% in all conditions. Due to the very few OS data-points, it was, therefore, not possible to perform statistical analyses for this group. Group 3 comprised the rest of participants. Analyses showed that the redundant use of colour adjectives was more frequent in the Colour-Reducing compared to both the Pattern-Reducing and Equally-Reducing conditions. Regarding the redundant use of pattern adjectives, however, none of the comparisons reached significance.

These results suggest that there is merit in all hypotheses regarding speakers' use of redundant adjectives, as different factors seem to motivate production strategies for different speakers. First, some participants adopted a strategy that minimised production effort by indiscriminately including redundant attributes in their utterances (Group 1). Even though these speakers did not always use both adjectives redundantly, their behaviour did not differ across conditions. This strategy matches the predictions of the egocentric view, since the use of redundant adjectives was not such that could facilitate listeners in their task (as shown in Experiments 2 and 3), but aimed mainly at easing production for the Speaker. Speakers in this group seem to have employed a simple strategy: They just started speaking quickly after the target referent was revealed, possibly without having fully scanned the scene for contrasting objects. Thus, they included in their utterances adjectives that turned out to be unnecessary (cf. Belke and Meyer, 2002; Pechmann, 1989).

The remaining participants seemed to take audience design into account when planning their utterance. A part of them almost never produced redundant adjectives, adopting a strategy that matches the predictions of the Gricean approach (Group 2). That is, in all conditions and independent of the entropy reduction potential of colour and pattern, they identified and uttered only the minimal information that was necessary, i.e., the object's type. The rest of the speakers (Group 3) used a strategy that took into account referential entropy, at least in the Colour-Reducing condition: OS rates for colour were higher in the Colour-Reducing condition compared to the Pattern-Reducing and Entropy-Reducing conditions. This result is in agreement with the bounded-rational approach, in that speakers used colour over-specifications rationally, i.e., more frequently when the colour adjective helped reduce entropy more effectively, contributing to a uniform reduction of uncertainty across the utterance. This tendency was observed only with redundant colour adjectives; redundant pattern adjectives were overall less frequent, which is, however, consistent with the findings from Experiments 2 and 3 showing that listeners prefer colour to pattern over-specifications.

Despite the vast diversity regarding what information speakers choose to use in their utterances, we were able to see that, at least for some of the speakers, referential entropy does play a role in whether and when to use a redundant adjective, in contexts where no adjective is actually needed to specify a referent. These speakers seemed to use over-specifications rationally, so as to guide their listeners' visual search for the intended referent. At the same time, however, they also satisfied their listeners general preference for colour over pattern over-specifications, as shown in the previous comprehension experiments.

5.2 Experiment 5

Experiment 5 examined the production of over-specified referring expressions in situations where the intended referent was part of a contrast pair, thereby requiring the mention of distinguishing information to be identified (cf. the blue striped ball in Fig. 5.3). Specifically, we evaluated whether the entropy reduction potential of a property (colour or pattern) in the referential space would influence speakers' redundant mention of this property. In other words, speakers may over-specify for a feature of the target referent not only because it stands out, but also based on the extent to which it reduces listener uncertainty about which object is the intended referent. For instance, speakers may be inclined to redundantly use an adjective such as 'blue' in order to identify a singleton object, not only because the colour blue is a salient property, and therefore easy to refer to, but also because it may help narrow down the referential space: If the set of blue objects is smaller than the set of other objects, the redundant mention of 'blue' would restrict the

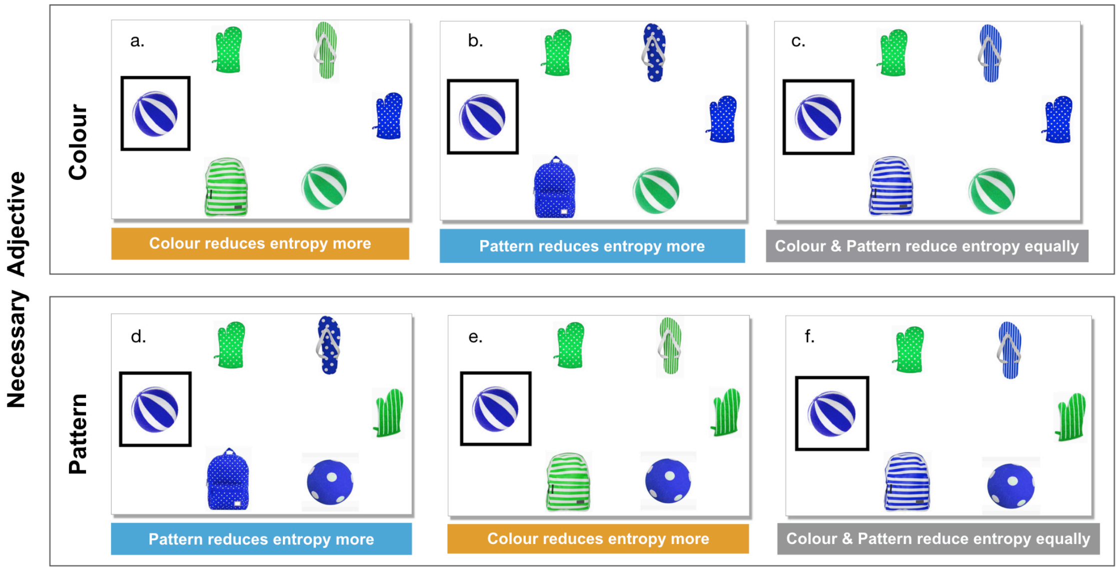


Figure 5.3 Experiment 5. Sample visual scenes (conditions) for an experimental item.

search space, and at the same time distribute the effort for target identification over a longer sequence of linguistic units. If, however, the blue objects outnumbered other objects, 'blue' would not be as effective as before in reducing uncertainty (the number of remaining referential candidates after hearing 'blue' would in this case be greater than before). Although a few recent studies have considered similar notions, such as discriminability, and their effects on referential over-specification (Fukumura, 2018; Koolen et al., 2016; Vogels et al., 2019), none of these studies directly controlled such factors. Thus, in Experiment 5 we manipulated which feature was Necessary for disambiguation (Colour-Necessary vs. Pattern-Necessary), and which feature was more Entropy-Reducing (Colour-Reducing vs. Pattern-Reducing vs. Equally-Reducing).

As in Experiment 4, the egocentric account predicted that speakers' OS rate should be independent of the experimental manipulation and not vary across conditions. On the other hand, according to the Gricean account speakers should prefer to convey the minimal amount of information in all conditions, as that would be expected by the listeners. That is, speakers should use the expression 'the blue ball' to refer to the intended referent in the Colour-Necessary conditions (cf. top panels in Fig. 5.3) and the expression 'the striped ball' in the Pattern-Necessary conditions (cf. bottom row in Fig. 5.3), independent of their entropy reduction potential. By contrast, the bounded-rational approach predicted that speakers should be more likely to over-specify particularly when the entropy reduction potential of the redundant adjective was higher than that of the necessary adjective. For example, in Figure 5.3e 'blue' would be redundant, but it would also reduce entropy at a higher extent compared to the necessary adjective 'striped' (ΔH_{blue} = 1.58 bits vs. $\Delta H_{striped}$ = 0.58 bits). Thus, the redundant 'blue' should be used more often in Figure 5.3e than in Figure 5.3d, where the necessary adjective ('striped') would be more entropy-reducing. Such production preferences would be in line with the findings from Experiments 2 and 3 that listeners favour utterances that reduce entropy more uniformly. Finally, as in Experiment 4 colour over-specifications were expected to be more frequent than pattern over-specifications.

5.2.1 Method

Participants. See Experiment 4.

Materials. The same object pictures as in Experiment 4 were utilised to create the visual stimuli. One experimental item comprised six versions of one display (six conditions; cf. Fig. 5.3), which differed in whether the mention of colour or pattern was required for disambiguation of the target referent (Colour-Necessary vs. Pattern-Necessary), and which of the features was more entropy-reducing compared to the other one (Colour-Reducing vs. Pattern-Reducing vs. Equally-Reducing). As in the previous experiments, critical displays contained six objects. The target referent was paired with another object of the same type (cf. the balls in Fig. 5.3), which differed from the target either in colour (cf. Fig. 5.3a–c) or in pattern (cf. Fig. 5.3d–f). A competitor object that shared the necessary feature with the target referent was included and was also part of a contrast pair (cf. the mitts in Fig. 5.3). This was done so that the use of a redundant adjective would not allow listeners to select the target immediately after hearing the adjective. Another two objects were included that differed in colour and pattern depending on the Entropy Reduction condition. That is, they differed from the target referent in the necessary feature, when this feature was more entropy-reducing than the other one (colour in Colour-Necessary, cf. Fig. 5.3a; pattern in Pattern-Necessary, cf. Fig. 5.3d); they shared the necessary feature with the target referent when the other feature was more entropy-reducing (pattern in

Colour-Necessary, Fig. 5.3b; colour in Pattern-Necessary, cf. Fig. 5.3e); they shared both features with the target referent when they were equally entropy-reducing (cf. Fig. 5.3c and f). A total of 216 displays were thus created making up 36 experimental items. Stimuli were intermixed with experimental and filler items from Experiment 4.

Procedure. See Experiment 4.

Data coding. Utterances were transcribed and annotated by the same annotators and following the same procedure as in Experiment 4. Data from two speakers who produced a high rate of under-specifications (more than 15%) were excluded from further analyses.[4] Data from one more speaker whose audio files were corrupted were not included in the analyses. Furthermore, faulty trials, that is trials in which speakers referred to the wrong object were excluded from further analyses (2.87%). Trials containing self-repairs on the adjective or the noun, and trials containing revisions of the utterance structure were excluded from analyses (4.2%). Over-specified utterances (57.05%) were coded as 1 and minimally-specified utterances (41.31%) as 0. A small number of trials containing under-specified utterances (1.64%; e.g., 'Is the ball on the left' in Fig. 5.3 a–c) were excluded from analyses. As in Experiment 4, in order to account for potential priming effects from the previous trial, after identifying speakers who regularly over-specified (16 participants, see below) we further excluded trials in which the same number of adjectives were used in the same word order as in the immediately preceding trial (11.93%).

Data analysis. Proportions of over-specifications were analysed using generalised linear mixed models with the lme4 package (Bates et al. 2015) in R (version 3.5.1; R Core Team 2018). The models included crossed random intercepts for both participants and items, and random slopes for the Necessary and the Entropy-reducing feature. Factors were treatment-coded, with Pattern as reference level for the Necessary feature, and Colour as reference level for the Entropy-reducing feature. When the maximal models did not converge, the random effects structure was simplified (Barr et al. 2013).

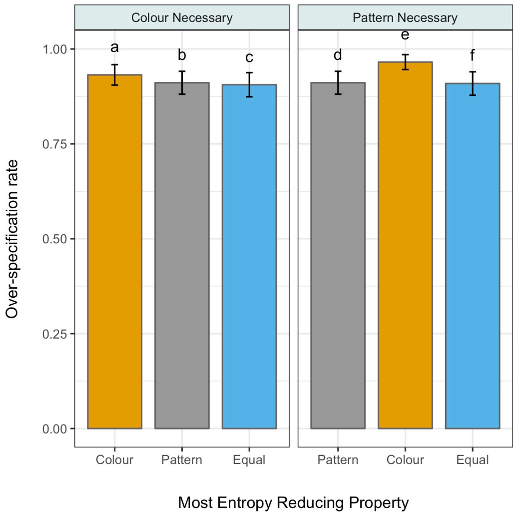
[4] A normal rate of under-specification is under 5% (cf. Davies & Katsos 2013; Engelhardt et al. 2006; Koolen et al. 2011; Koolen et al. 2013; Koolen et al. 2015, among others).

5.2.2 Results

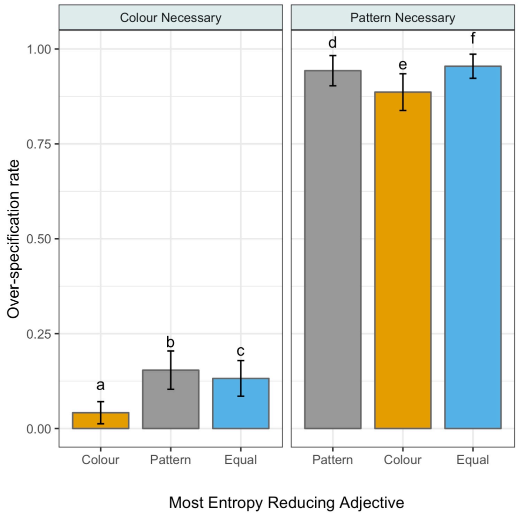
The egocentric view predicted that Speakers' OS rates should not be influenced by our manipulation. Indeed, 16 participants were found to over-specify in most of the trials. Interestingly another 10 participants over-specified regularly, but only for colour. Based on this pattern of results, we categorised participants into three groups depending on their general pattern of OS use.[5] Group 1 included speakers who produced both adjectives more than 80% of the time (N = 16; see Fig. 5.4a), Group 2 consisted of speakers who produced redundant colour adjectives more than 80% of the time (N = 10; Fig. 5.4b), and Group 3 consisted of the remaining 16 participants (see Fig. 5.4c). Analyses were performed per group.

Results from all groups are summarised in Table 5.1. In Group 1 none of the comparisons reached significance. Over-specifications were equally frequent for both colour and pattern in all conditions (Fig. 5.4a). In Group 2, only the Necessary feature was found to be significant, with more over-specifications in the Pattern-Necessary than in the Colour-Necessary conditions (i.e., redundant colour adjectives were used more frequently than redundant pattern adjectives; cf. Fig. 5.4b). In Group 3, the Necessary feature resulted in a significant effect, with more over-specifications in the Pattern-Necessary than in the Colour-Necessary condition (i.e., more over-specifications for colour than for pattern), but further comparisons were also found to be significant. In particular, regarding the Entropy-reducing factor, the comparison between Colour-Reducing (reference level) and Pattern-Reducing yielded a marginally significant effect indicating a higher proportion of OS when colour reduced entropy more than pattern (cf. yellow and grey bars, respectively, in Fig. 5.4c). Moreover two Necessary x Entropy-Reducing interactions were found (see Table 5.1). The difference in OS rate between Pattern-Necessary and Colour-Necessary was larger for the Colour-Reducing than the Pattern-Reducing conditions (cf. the difference between yellow and grey bars in the two panels of Fig. 5.4c). Similarly, the difference in OS rate between Pattern-Necessary and Colour-Necessary was larger for the Colour-Reducing compared to the Equally-Reducing conditions (cf. the difference between yellow and blue bars in the two panels of Fig. 5.4c).

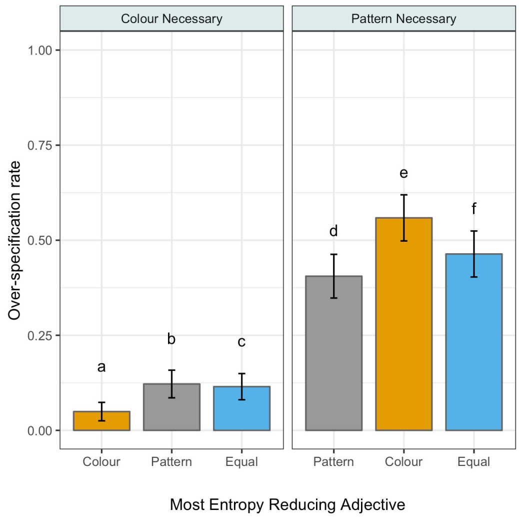
[5]Three of the originally 48 speakers produced minimal descriptions more than 90% of the time. While these participants can be considered as highly Gricean, the current study was concerned with understanding why speakers over-specify. Therefore, no further analyses were pursued for their data.



(a) Group 1 (N = 16)



(b) Group 2 (N = 10)



(c) Group 3 (N = 16)

Figure 5.4 Experiment 5. Proportions of over-specifications produced by each group of participants. Error bars represent SE.

Table 5.1 Experiment 5. Coefficients, *SE*, *z*- and *p*- values for the comparisons of over-specification rates between the reference levels (Pattern-Necessary and Colour-Reducing) and the other levels of the fixed factor, and their interactions.

Effect	Group 1				Group 2				Group 3			
	β	*SE*	*z*	*p*	β	*SE*	*z*	*p*	β	*SE*	*z*	*p*
Intercept	3.862	0.702	5.498	<.001***	2.588	0.677	3.822	<.001***	0.255	0.274	0.929	.353
Necessary												
Colour	-0.747	0.763	-0.978	.328	-6.159	1.136	-5.422	<.001***	-3.266	0.574	-5.687	<.001***
Entropy-Reducing												
Pattern	-1.115	0.734	-1.517	.129	0.896	0.922	0.973	.33	-0.65	0.349	-1.86	.0629 .
Equal	-1.139	0.735	-1.549	.121	0.994	0.91	1.092	.275	-0.374	0.352	-1.064	.287
Necessary:Entropy-Reducing												
Colour-N:Pattern-R	0.856	0.962	0.889	.374	0.577	1.262	0.457	.647	1.618	0.708	2.284	.0224 *
Colour-N:Equal-R	0.685	0.953	0.719	.472	0.34	1.266	0.269	.788	1.274	0.71	1.796	.0726 .

Note: Colour-N = Colour-Necessary, Pattern-R = Pattern-Reducing, Equal-R = Equally-Reducing

5.2.3 Discussion

In Experiment 5, we investigated whether the type of feature that is necessary to specify a target object from within a contrast pair, as well as the relative entropy reduction potential of a necessary vs. redundant feature influences the production of over-specifications. In a referential communication task, we manipulated whether colour or pattern was necessary to identify the target referent, and which of these features was more Entropy-Reducing (Colour-Reducing vs. Pattern-Reducing vs. Equally-Reducing). We measured participants' over-specification rate in each condition. The higher overall rate of minimally-specified referring expressions (59%) compared to over-specified referring expressions (39%) clearly demonstrates that speakers are able to produce and often do produce minimal descriptions. What we were interested in understanding, however, is under what circumstances they over-specify. As in Experiment 4, we found that speakers adopted different production strategies. We, therefore, split participants in groups according to their general pattern of over-specification. We first identified a group (Group 1) that over-specified more than 80% of the time with both pattern (i.e., in the Colour-Necessary conditions) and colour (i.e., in the Pattern-Necessary conditions) adjectives. That is, these participants very rarely produced expressions that did not encode both adjectives. This behaviour is in line with the predictions from the egocentric view, that speakers' use of over-specification should not be affected by our experimental manipulation. This result indicates that at least for some speakers, over-specification is a way to (safely) lift some of the target specification burden, e.g., by always using a template that contains both modifiers, regardless of the visual environment.

Group 2 included participants whose over-specification rate was greater than 80%, but only in the Pattern-Necessary conditions. That is, they regularly used a redundant colour adjective, regardless of its entropy reduction potential. The results from this group are in accord with both the egocentric view and with the audience-design view in that speakers prefer to use redundant colour adjectives more frequently than redundant pattern adjectives. On the one hand, this preference may be due to colour salience, which eases property selection for the speakers. On the other hand, it also facilitates target identification for the listeners, who favour colour over-specifications over pattern over-specifications, as was shown in our comprehension experiments.

Finally, the remainder of participants were grouped together (Group 3). Results showed that speakers in this group also over-specified more for colour than for pattern, but their use of redundant adjectives varied with the distributional properties of the visual scene: They over-specified more frequently when the redundant adjective was more entropy-reducing than the necessary adjective. This behaviour matches the predictions of the bounded-rational approach, which argues that speakers should over-specify more when the redundant adjective reduces referential entropy to a higher degree than the necessary adjective. These results also fit the findings from Experiments 2 and 3, which showed that listeners favour over-specified expressions, as well as expressions that reduce entropy at a high rate early on, but this preference is greater for colour than for pattern adjectives.

Overall, while individual differences seem to govern production choices, we found evidence for the use of a bounded-rational strategy (Group 3). This strategy appears to take into account the distributional properties of the visual scene in order to ease the listener's task, by producing a redundant adjective more frequently when it helps reduce entropy at a higher rate than the necessary adjective. We further found that egocentric concerns may also be at play in referential communication, and that at least in some cases (Group 1) over-specifications may be 'for the speaker'. A third strategy was also observed (Group 2), in which over-specifications were used independent of condition, but only with colour adjectives. This strategy could be interpreted either under an audience-design or an egocentric view, as colour is a visually salient property and arguably preferred by both speakers and listeners in such tasks. Evidence in support of the Gricean account in this experiment was minimal; only three of our participants systematically used minimal information in all conditions.

5.3 Summary

Experiments 4 and 5 aimed at evaluating whether the factors that were previously found to influence comprehension (i.e., target feature and entropy reduction on the adjective) may motivate speakers' use of over-specification. In both experiments, we manipulated the factor that was most Entropy-Reducing (Colour-Reducing vs. Pattern-Reducing vs. Equally-Reducing). The two experiments differed in whether the target referent was a singleton object, which did not require modification to be specified (Experiment 4), or it was part of a contrast pair, thereby requiring modification (Experiment 5). In this last case, we further manipulated the kind of adjective that was Necessary (colour vs. pattern).

According to the *Gricean* account, speakers should encode only the minimal amount of information that is every time necessary to specify the target referent. Nevertheless, as it has repeatedly been shown in the past (see Section 2.1.1), speakers frequently over-specify, and it is this behaviour that we seek to explain here. In principle, two forces may motivate the production of redundant adjectives. On the one hand, over-specifications may be the result of a strategy to ease production-internal processes. Under this *egocentric* hypothesis speakers may start planning and executing an utterance before they have fully scanned the visual scene, that is before they have identified what information is necessary to specify the target referent. Thus, they include adjectives (relying on heuristics based on their prior experience) that turn out to be redundant. On the other hand, speakers' production of over-specifications may be tuned to facilitating the listeners in their task. More specifically, as we saw in the previous comprehension experiments, the use of an adjective that reduces referential entropy at a higher rate contributing to a more uniform entropy reduction profile benefits listeners' visual search for the target object, while it also boosts their performance in the task, even if this adjective is redundant (see P600 and RTs in Exp. 2, ICA and RTs in Exp. 3). This effect was found to hold more for colour than for pattern adjectives. Therefore, under this *bounded-rational* hypothesis speakers are rational agents and their over-specifications should be motivated by a desire to attend to the listeners' needs and, thereby, improve communication.

In both experiments, we found that speakers' utterances were marked by a high degree of diversity, highlighting individual differences in language production. In particular, in both experiments speakers adopted different production strategies, and in each experiment we grouped them based on the overall strategy that they adopted. In Experiment 4, a first group systematically over-specified with both colour and pattern adjectives (Group 1). A second group consisted of speakers who almost never over-specified (Group 2). Group 3 comprised the remaining participants, who were found to produce colour adjectives more frequently when they had a high entropy-reduction potential. The strategies we identified in Experiment 5, were quite similar. One group of speakers regularly used both adjectives regardless of which one was required (Group 1). A second group also over-specified frequently, but only for colour (Group 2). The remaining participants were categorised in the last group (Group 3), and were found to produce redundant colour adjectives more frequently when these adjectives reduced entropy at a higher rate than the necessary pattern adjectives. In Experiment 5, only three speakers were found to regularly use minimal descriptions.

Overall, even though variation in speakers' production choices was quite high, there is a systematic pattern in our results that gives relative merit to all approaches of referential communication. That is, in both experiments a group of speakers over-specified very frequently regardless of the visual context (Group 1 in both experiments). These speakers adopted a strategy that eased production processes for themselves, in line with the egocentric view. Furthermore, in both experiments we identified speakers who were highly Gricean, i.e., they almost never used unnecessary adjectives. Speakers who adopted this strategy were more in Experiment 4 (13 speakers of Group 2) than in Experiment 5 (3 speakers). This discrepancy may indicate that it is more likely for speakers to include a redundant adjective in their utterances, when one adjective is already required. In both experiments, some of the speakers took into consideration the complexity of the visual scene and tried to modulate it in their utterances (Group 3 in both experiments). Both with singletons (Experiment 4) and with contrasting target objects (Experiment 5), these speakers seemed to consider the (relative) entropy-reduction potential of the redundant adjective when planning their utterances. What is more, this behaviour was more likely with colour than with pattern over-specifications. Therefore, this behaviour confirms the bounded-rational approach, but is also in line with the findings from Experiments 2 and 3, that listeners prefer redundant colour over redundant pattern adjectives. Lastly, in Experiment 5 only, a group of speakers was found to over-specify regularly but only for colour (Group 2). This strategy was not anticipated, but is in line with the predictions of both the egocentric view and the audience-design view, as our comprehension experiments show a general preference for colour over pattern over-specifications.

To summarise, in two production experiments we have shown (a) that speakers' production preferences are highly variable, offering some support to all of the hypotheses that we considered, and (b) that, at least for some speakers, the relative entropy-reduction advantage of adjectives may motivate their redundant inclusion in referring expressions.

Chapter 6

A Bounded-Rational Account of Referential Redundancy

Referential redundancy poses a puzzle for Gricean pragmatic accounts of communication, according to which, speakers who intend to specify a target referent in a visual scene should produce only the information that is necessary and nothing more. As the addressees expect speakers to conform to such conversational norms, their comprehension should be disrupted by redundant information. In other words, rational speakers should always choose minimal descriptions in order to specify an intended referent, and rational listeners should always expect that the information they receive is necessary for this task. Empirical studies have, however, time and again shown that speakers actually encode unnecessary information in their referring expressions quite frequently, which raises the question: *Why do speakers over-specify?*

There are, generally speaking, two approaches to answering this question. According to the *egocentric* approach, speakers mainly aim at reducing production effort, which may lead them to start speaking before they have exhaustively scanned the visual scene for possible competitor objects to the target. As a result, they might include information that ends up being unnecessary. According to the *audience-design* approach, on the other hand, speakers' over-specifications actually aim at facilitating referential processing for their addressees. Deciding between these accounts is, therefore, heavily dependent on how addressees respond to over-specifications. Evidence on the comprehension of over-specifications is, however, mixed. Some studies find that they hinder comprehension while others show that they might in fact facilitate it.

This thesis explored whether a *bounded-rational account* of communication is better suited to explain the use of redundancy, reconciling the seemingly incompatible empirical evidence. We propose that rational speakers should strive to *minimise joint effort* in establishing reference, by producing redundant descriptions when the extra information helps to reduce uncertainty about the target referent more efficiently (i.e., fast, easily and reliably) for their listeners. In other words, minimal descriptions may not always be optimal; depending on factors such as the complexity of the visual scene, the importance of the task, etc., redundant information may benefit referential communication.

Recent communication accounts that are based on Information Theory (Shannon, 1948) have shown that we can measure the informativity of a word as the expectancy of this word in its context (surprisal; see Hale, 2001; Levy, 2008), or as the amount of reduction in uncertainty about the rest of the sentence (entropy) that this word induces (see Hale, 2003, 2006). These information values have been associated with processing effort: The higher the information in bits on a given word (higher surprisal or reduction of entropy), the more cognitive effort is required for processing that word (cf. Frank, 2013; Smith and Levy, 2013). The Uniform Information Density (UID) hypothesis then predicts that speakers' production choices may be determined by an interest to distribute information (and thereby, cognitive effort) across more linguistic units (Jaeger, 2010; Levy and Jaeger, 2007). That is, when faced with two meaning-preserving syntactic alternatives to express a message, speakers are more likely to choose the one that distributes information more uniformly across the utterance, in order to avoid peaks in cognitive load that may exceed the bounded processing resources of their listeners. For example, speakers may consider a verb's subcategorisation preferences when planning a complement clause: In case a complement clause (e.g., 'we were absolutely crazy') is not highly likely after the main verb (e.g., 'confirmed'), speakers may choose to insert the optional *that* complementizer. Otherwise the first word in the complement clause (e.g., 'we') would be ambiguous between the onset of the clause and its semantic content, and thus highly informative. Speakers may, therefore, choose encodings that distribute information content across more linguistic units (and time), thus avoiding informationally dense peaks in their utterances causing processing difficulty for the listeners.

Existing accounts measure informativity as determined by linguistic contexts. We extended these accounts into visually-situated communication, where the informativity of a word is determined based on its probability to occur both in the current utterance and in the particular visual scene. We introduced the notion of *referential entropy* as a measure of the uncertainty about which object in a visual scene will be mentioned in the referring expression. For instance, referential entropy at '*blue*' in 'Find the blue. . . ' is lower in a visual scene containing two blue objects compared to a scene containing four blue objects, even though in both cases the utterance is exactly the same. That is, '*blue*' reduces referential entropy to a different extent depending on the availability of blue objects in the visual context. By extension of the Entropy Reduction Hypothesis (Hale, 2003, 2006) into the visual world (i.e., measuring the reduction of entropy over potential referents in a visual scene instead of possible sentence continuations), the cognitive effort for processing a word should be proportional to the degree of referential entropy reduction induced by that word, in bits of information. In order to successfully establish reference, referential entropy must, of course, be reduced to zero, and speakers can formulate their utterances so as to modulate the rate at which the initial entropy is minimised.

We, therefore, propose a bounded-rational account of referential redundancy, according to which, any word in an utterance, even if it is redundant, will be beneficial to comprehension, as long as it helps reduce referential entropy. Under this account, over-specifications may be utilised by speakers in order to distribute the effort associated with processing the referring expression and identifying the target referent, more uniformly across the expression. In order to assess this account, we developed three research questions, each building on the answer to the previous one, and designed a set of comprehension and production experiments addressing each of them. Our research questions were:

- RQ-1 Are over-specifications detrimental or beneficial to comprehension?
- RQ-2 Is the comprehension of referring expressions influenced by how efficiently they manage the reduction of referential entropy?
- RQ-3 Are speakers sensitive to the distributional properties of the visual scene, and if so do they use over-specifications as a means to modulate referential entropy for their listeners?

6.1 Summary of Experimental Results

A first ERP experiment addressed RQ-1. Participants were presented with complex visual scenes, depicting six objects that differed along three dimensions (colour, pattern and type), paired with one audio instruction to locate one of the objects in the scene. We manipulated the Specificity of the referring expression: While the audio instruction (e.g., 'Find the yellow bowl') was held constant across conditions, the visual scenes differed, rendering the utterance minimally-specified (MS), over-specified (OS) or under-specified (US). In the MS condition the visual scene depicted two objects of the same type (e.g., two bowls) differing in one feature (colour or pattern), and the utterance, therefore, carried the minimal information for identifying the target referent. In the OS condition, only one object of the mentioned type was available in the visual scene (e.g., one bowl), and, therefore, the adjective in the referring expression was redundant. In both the MS and OS conditions, however, reference could be successfully established; it was unambiguous which object was the target referent. The US condition, on the other hand, resulted in referential failure: Even though the utterance identified the type of the target referent, it did not help listeners disambiguate between the two same-type objects available in the scene (e.g., two yellow bowls). In order to establish whether referential failure due to under-specification is similar to *explicit* referential failure due to a mismatch between the visual and linguistic input, a Mismatch (MM) condition was also used. In this condition, the adjective predicted a single referent (i.e., there was only one yellow object in the visual scene), which was, however, not the one mentioned by the noun.

Our results presented two important insights. First, our results distinguished two qualitatively different processes associated with referential failure: referential failure due to the lack of information, and referential failure due to the mismatch between linguistic and visual information. Specifically, the US condition yielded a long-lasting positivity compared to the MS condition, starting around 400 ms after the onset of the adjective and sustained throughout the noun time-window. The MM condition, on the other hand, elicited a large N400 effect in the noun time-window (see below).

Secondly – and more central to the concerns of this thesis – in complex visual contexts over-specifications were found to benefit comprehension. Specifically, we observed a graded N400 effect for the MS, OS and MM conditions in the noun region, with the OS condition eliciting the most reduced N400 amplitude and the MM condition eliciting the highest N400 amplitude. Given that the N400 component has been linked to word expectancy, we interpreted this graded N400 effect as evidence of the visually-determined surprisal on the noun. That is, while in the MM condition the head noun (e.g., 'watering can') was unexpected based on the preceding adjective (e.g., 'yellow'), in the OS condition the referent mentioned by the noun (e.g., bowl) was the only referent in the scene that matched the adjective, and therefore it was highly predictable, compared to the MS condition, where the adjective selected two referents. This finding comes in contrast to a strict rational (Gricean) account of communication, but is in line with our bounded-rational account.

Experiments 2 and 3 investigated whether the rate at which referential entropy is reduced across a referring expression also affects processing, and whether Entropy Reduction and Specificity have additive effects (RQ-2). That is, these experiments investigated whether expressions that reduce referential entropy more efficiently (i.e., distribute the reduction of referential entropy more evenly across the utterance) benefit processing, and whether effects of Specificity may be observed above and beyond any effects of Entropy Reduction. It is, however, possible that the facilitation observed for over-specifications compared to minimal descriptions in Experiment 1 was a by-product of the display structure in the OS condition of that experiment, where only one object matched the adjective and this object was a singleton. That is, the visual scenes in the OS condition supported only a non-contrastive interpretation of the adjective. This issue was also addressed in Experiments 2 and 3, by using visual scenes that supported both a contrastive and a non-contrastive reading of the adjective; i.e., both referents in contrast pairs and singleton referents matched the adjective and were simultaneously available in the scenes.

We created a set of stimuli, all of which were used in an ERP experiment (Exp. 2) and half were used in an eye-tracking experiment (Exp. 3). In both experiments, we, furthermore, measured the Index of Cognitive Activity (ICA), which is a pupillary measure and can be recorded during both ERP and eye-tracking experiments, as an index of overall cognitive effort. We expected that results on the ICA would be similar in the two experiments, thus making it possible to correlate the ERP with the eye-tracking results. Four visual displays were paired with one spoken instruction. The instruction was minimally-specified (MS) when it identified one object from within a contrast pair, and was over-specified (OS) when it identified a singleton object. Furthermore, referential entropy was reduced at a higher rate (HR) when the prenominal adjective selected two potential referents, while it was reduced at a lower rate (LR) when the adjective selected four referents. Four experimental conditions were generated in this way: MS-HR, MS-LR, OS-HR, and OS-LR.

The results from these experiments paint a complex picture. First, we found evidence that over-specification affects comprehension in different ways depending on the nature of the redundant adjective used in the expression. Redundant pattern adjectives were incrementally interpreted contrastively (more anticipatory eye movements to the contrast vs. singleton objects during the adjective), which hindered processing on the following noun (increased N400 for the OS vs. MS conditions). On the other hand, colour adjectives were not found to be assigned a contrastive reading (no difference in anticipatory eye movements to contrast vs. singleton objects during the adjective) nor did they negatively affect processing of the head noun (no difference in the N400 between the OS and MS conditions).

At the same time, both kinds of redundant adjectives (colour and pattern) were found to facilitate the listeners' *visual search* for identifying the target referent and performing the task. That is, even when on-line comprehension was hindered by redundancy, the redundant adjective offered additional visual cues allowing listeners to restrict the set of potential referents and easing their task performance (indicate which side of the screen the target referent appeared on). This interpretation is supported by ERP (P600 effect for the OS conditions), eye-tracking (more fixations to the target vs. competitor object in the OS vs. MS conditions), and pupillary (lower ICA values in the OS vs. MS conditions during the noun) data in the noun region, as well as by behavioural evidence (faster RTs in the OS vs. MS conditions). These effects were observed for both colour and pattern redundant adjectives, but were stronger for the colour adjectives.

Furthermore, our results showed that the rate at which referential entropy is reduced during the utterance also influences comprehension processes. In line with the Entropy Reduction Hypothesis (Hale, 2003, 2006), a high rate of entropy reduction on the (redundant or necessary) adjective (i.e., an adjective selecting 2 out of 6 referents, compared to 4 out of 6 referents) resulted in greater cognitive effort in that region, but lower effort on the following noun. Interestingly, this was an effect observed only in the ICA measures of Experiment 3 (i.e., only in the eye-tracking and not in the ERP experiment). As the only difference between the two experiments was that in the ERP experiment participants could not freely scan the visual scene, while in the eye-tracking experiment they could, we believe that the cognitive effort that the ICA is sensitive to stems from the difficulty of the visually-grounded identification of the target referent. In other words, the ICA captures the cognitive effort of scanning the visual scene in order to narrow down the set of potential referents and identify the location of the target. This difficulty was higher on the adjective but lower on the noun for the HR compared to the LR conditions (recall that the effects of Entropy Reduction on the ICA were similar for both colour and pattern items). In addition to the effects on the ICA, we also found that a high reduction of entropy on the (redundant or necessary) adjective facilitated processing on the subsequent noun (reduced N400 effect for HR vs. LR) and improved participants' target identification times (faster RTs in HR vs. LR). An interaction between Specificity and Entropy Reduction was also observed, but only in one of our measurements (ERPs) and only with colour adjectives. That is, redundant colour adjectives were found to particularly enhance participants' task performance, when they resulted in a higher reduction of referential entropy (P600 effect for the OS-HR condition on the noun). Interestingly, using ICA we were able to directly compare the cognitive effort for processing over-specified utterances (e.g., 'Find the blue ball' in a visual scene with only one ball; OS-HR condition) to that for their minimally-specified counterparts (e.g., 'Find the ball' in the same scene; MS fillers). This comparison suggested that listeners' cognitive effort for identifying the target referent was lower for nouns that followed redundant adjectives (OS-HR condition) compared to bare nouns (MS fillers).

Summarising, our comprehension experiments showed that, depending on the visual context, over-specifications may facilitate processing. In visual contexts that support both a contrastive and a non-contrastive interpretation of the adjective (i.e., where at least two objects matched the adjective, one in a contrast pair and one singleton), on-line comprehension was hindered by redundant pattern adjectives, but not by redundant colour adjectives (anticipatory looks to singleton vs. contrast objects on the adjective, and N400 on the noun). This finding is in line with both the Gricean (pattern adjectives) and the bounded-rational (colour adjectives) accounts. However, the listeners' visual search for the target referent was benefited by referential redundancy *regardless* of the nature of the adjective used (P600 and ICA on the noun, overall RTs), supporting the bounded-rational account. In any case, the rate at which referential entropy is reduced during the utterance was found to further influence comprehension. A high rate of entropy reduction on the adjective resulted in difficulty in that region, but facilitated the on-line comprehension and the visual grounding of the referring expression on the following noun (reduced

N400, ICA, RTs). Lastly, redundant colour adjectives inducing a high rate of entropy reduction particularly benefited participants' task performance (P600 for the OS-HR condition on the noun).

After establishing an understanding of how over-specifications are processed online, and how referential entropy reduction further affects processing, we moved to the production end of referential communication, and tested whether such comprehension preferences are taken into consideration by the speakers. In two production experiments, we examined speakers' use of over-specifications in visual contexts where no adjective was necessary to identify the target referent (Exp. 4), and in contexts where only one adjective was required (Exp. 5). In both experiments, we manipulated the entropy reduction potential of a redundant adjective: whether a redundant colour adjective would reduce referential entropy more or less compared to a redundant pattern adjective, or they would both reduce entropy to an equal extent. This yielded three experimental conditions: Colour-Reducing, Pattern-Reducing and Equally-Reducing. In Experiment 5, these conditions were crossed with the Necessary Adjective manipulation, i.e., whether a colour or a pattern adjective was necessary for identifying the target referent.

We generally expected our results to be in line with either the egocentric or the audience-design views. According to the egocentric view, speakers over-specify in order to ease production-internal processes. Therefore, speakers' referential choices are not expected to be influenced by the distribution of objects in the visual scene, and their OS rate should not vary across conditions. The Gricean and the bounded-rational accounts both fall under the audience-design view. Based on the Gricean account, speakers should produce only minimal descriptions, because that is what their listeners expect, and over-specifications would, therefore, disrupt comprehension. Based on the bounded-rational account, on the other hand, speakers are expected to use redundant adjectives when they have a generally high entropy-reduction potential (Exp. 4), or when they are more entropy-reducing than the necessary adjective (Exp. 5). Additionally, based on the results from our comprehension studies, we expected that the OS rate for colour adjectives should be overall higher than the OS rate for pattern adjectives, as listeners were found to have a preference for colour over pattern redundant adjectives.

The results offer two important insights regarding the production of over-specifications. First, that speakers' production choices are highly diverse; in both experiments we identified groups of speakers opting for different production strategies. Secondly, we found that, at least for some of the speakers, referential choices may be guided by a concern to efficiently reduce entropy across their utterances, supporting the bounded-rational account. More specifically, in both experiments we categorised participants in groups, based on their production choices. In Experiment 4, a first group of speakers produced over-specifications with either colour, pattern, or both adjectives more than 80% of the time in all conditions, in line with the egocentric view. A second group of speakers produced almost exclusively (more than 90%) minimal descriptions in all conditions. These speakers adopted a Gricean production strategy. The rest of the speakers were grouped together, and were found to over-specify more frequently when the redundant adjective had a high entropy-reduction potential, in line with the bounded-rational account.

In a similar vein, in Experiment 5 we identified a first group of speakers who produced redundant adjectives more than 80% of the time regardless of the condition, as the egocentric account predicts. A second group was formed by speakers who over-specified for colour more than 80% of the time. This is a strategy which is in line with both the egocentric and the bounded-rational accounts: Colour is a salient feature, and attracts speakers' attention, who are inclined to start their utterance with it, and possibly buy extra time to identify which other features – if any – they need to mention. At the same time, because of its salience, colour can rapidly facilitate listeners' visual search for the target, as our comprehension experiments showed. Lastly, the remainder of the speakers were grouped together, and were found to over-specify more frequently when the redundant adjective was more entropy-reducing than the necessary adjective. This strategy was modulated by the well-established by now preference for colour over pattern adjectives, and is in line with the bounded-rational account.

To summarise, the results from two production experiments provide support for both egocentric and audience-design views on referential redundancy. Crucially, a large number of speakers in both experiments adopted a rational over-specification strategy: They over-specified more frequently when the redundant adjective helped them manage referential entropy more efficiently.

6.2 Key contributions

As highlighted in the introduction, referential redundancy poses a problem to theories of communication that assume speakers and listeners to be (strictly) rational, such as the Gricean theory. Many studies have asked how redundancy affects comprehension for the listeners and why it is used by speakers, offering mixed results. In our research, we sought to shed light on this debate, and offer a unified explanation for the production and comprehension of over-specifications. In this effort, we conducted both comprehension and production experiments combining different experimental paradigms and measures. Therefore, apart from theoretical value, our results have additional implications regarding the experimental methodologies that we used. In what follows, we discuss the theoretical and methodological contributions of our work.

6.2.1 Theoretical contributions

In recent years, Information Theory has been re-introduced in linguistic research as a framework for quantifying complexity and explaining production and comprehension preferences. Information-theoretic metrics, such as surprisal and entropy reduction, have been shown to provide estimates of the cognitive effort associated with processing a word in its context. Prior work has, however, focused on how these metrics are modulated by the linguistic context. Surprisal, therefore, provides an estimate of word predictability based on the words that appeared previously in the sentence (Hale, 2001; Levy, 2008), while entropy reduction measures the amount of reduction in uncertainty over the possible syntactic continuations of the sentence (Hale, 2003, 2006).

In our studies, we extended these information-theoretic notions into visually-situated communication. We paired visual scenes with spoken descriptions of one of the objects in the scene, and calculated entropy reduction based on the number of referents that incrementally matched the referring expression; i.e., the number of objects that would potentially be referred to at each word as the expression unfolded over time (see Venhuizen et al., 2019, regarding models of semantic entropy reduction). Crucially, in the visual stimuli we used only images of artificial objects (e.g., balls, mugs, etc.) differing for their colour and pattern, thereby ensuring that (a) no specific noun would be predicted after a colour adjective, because none of our referents had a natural colour (cf. yellow banana vs. yellow bowl), and (b) that all nouns represented by referents in one scene would be equally predictable after any adjective (e.g., a bowl and a watering-can were equally likely to be yellow). In this way, we made sure that only the distributional properties of the visual context would determine word predictability, as the linguistic context was neutral.

We used referential entropy as a measure of uncertainty regarding the target referent: the higher the number of referents selected by a word in the utterance, the higher the uncertainty regarding the target referent at that point. By extending the Entropy Reduction Hypothesis (Hale, 2003, 2006) in visual contexts, we, therefore, expected processing effort on a word to be proportional to the amount of *reduction* in referential entropy induced by that word. And by extension of the Uniform Information Density hypothesis (Jaeger, 2010; Levy and Jaeger, 2007), we expected that speakers would seek to distribute the effort associated with referential entropy reduction across more linguistic units, thereby inserting redundant words in their utterances.

Our experiments generally showed that, when communication takes place in the presence of a (relevant) visual context, the structure of that context influences word expectancy, potentially overriding gricean considerations, i.e., that adjectives should be used contrastively. Our first experiment found effects of referential surprisal, although we did not explicitly manipulate it as a factor. When a redundant adjective narrowed down the set of potential referents to a single object, listeners utilised this information to predict the upcoming noun, without ruminating on the redundancy of the adjective. Redundant adjectives were, therefore, found to facilitate processing of the noun they modified, because they increased the expectancy of that noun. This visually-determined surprisal modulated the N400 component, similarly to word surprisal in linguistic contexts. Our other experiments directly manipulated Entropy Reduction and found that the rate at which uncertainty about the target referent is decreased across the referring expression does indeed affect referential processing, and can motivate the use of a redundant adjective, at least for some speakers.

The use of different methodologies testing the same experimental material in Experiments 2 and 3 yielded interesting results that tap into two distinct comprehension processes in referential communication: (a) the (visually-grounded) comprehension of the referring expression, and (b) the visual search for identifying the target referent and performing the task. (See more on the methodological aspects of identifying this distinction in the following Section). Our results suggest that referential redundancy may affect these processes in different ways, depending on the kind of redundant information used. That is, redundant pattern – but not colour – adjectives hinder the on-line interpretation of the expression, while at the same time both kinds of adjectives facilitate the visual identification of the target referent and thus participants' performance in the task (i.e., pressing a button to indicate the target object's location on the screen).

Lastly, our work offers a bounded-rational account of referential over-specification, unifying evidence from comprehension and production studies. According to this account listeners' on-line comprehension and target identification processes is benefited by the use of a redundant adjective, when this adjective helps reduce uncertainty about what is being communicated. Rational speakers may, therefore, encode redundant adjectives in their referring expressions, when the communicative situation is challenging, in order to ease speakers' and listeners' joint effort in achieving their shared goal, i.e., to establish reference (see Section 6.3.2).

6.2.2 Methodological contributions

Apart from contributions on the theoretical level, our work has offered important insights regarding the use of different methodologies and the interpretation of the results acquired with these methodologies.

Our experiments were among the first to use the ICA as a measure of cognitive load in language processing. Even though pupillometry has been long used in psycholinguistics to assess cognitive effort (see Just and Carpenter, 1993), the ICA is a technique that was developed in 2000 and was only recently employed in linguistic research (see Demberg and Sayeed, 2016). Furthermore, only a handful of studies have used the ICA in combination with the visual world paradigm (Ankener et al., 2018; Sekicki and Staudte, 2018; Vogels et al., 2018), where participants were allowed to move their eyes across the visual scene as they listened to spoken linguistic stimuli. Because in our research we recorded the ICA in combination with EEG on a set of stimuli that was also tested with eye-tracking/ICA alone, we were able to directly compare the results of the two experiments and draw interesting conclusions regarding the kinds of processes that the different methodologies tap into. Specifically, due to the nature of EEG experimentation, in the EEG/ICA experiment (Exp. 2) participants had to fixate a point in the middle of the screen as they listened to the spoken instruction. By contrast, in the eye-tracking/ICA experiment (Exp. 3), participants were allowed to freely scan the visual scene as the audio stimulus was played. What we observed was that in Experiment 2, participants' ICA values sharply decreased soon after the start of the spoken instruction (at which point participants should be fixating in the middle of the screen), and remained flat throughout the utterance. In Experiment 3, on the other hand, we found that ICA values started off low, and incrementally increased at a different rate depending on the condition. From this comparison, we were, therefore, able to infer that the kind of cognitive workload indexed by the ICA is related to *visually-grounded* linguistic processing. That is, we suggest that in visually-situated comprehension the ICA indexes the effort involved in performing the task of identifying the location of the target referent, and is suppressed when *overt* visual search is impossible.

Interestingly, as already mentioned in Section 6.2.1 above, in the Pattern items the results acquired with the ICA did not match the ERP results, pointing to different aspects of referential processing. Although the ICA values were lower on the noun after a redundant pattern adjective, indicating that cognitive load was decreased in that region, the ERP results were suggestive of a processing difficulty in the same region (increased N400 amplitude). Because (a) the ICA was shown to be associated with visually-grounded linguistic processing, and (b) the ICA results matched with a task-related facilitation observed with other measures (RTs, P600), we believe that the ICA is sensitive to visual processes for identifying the target referent. The difficulty found in the N400 results, on the other hand, independently indexes on-line comprehension processes.

Lastly, our results speak to the hypothesis that relates attention to linguistic processing in visual-world eye-tracking. According to this *linking hypothesis*, increased visual attention is evidence of facilitated processing (cf. Tanenhaus et al., 2000). Visual-world studies, therefore, predict that segments in the linguistic input that are easily processed should evoke more fixations to the related area of the visual scene, while processing difficulty should manifest as a lower fixation rate in that area. In colour items of Experiment 3, results on the noun showed a higher probability of fixating the target over the competitor object in the HR vs. LR conditions and in the OS vs. MS conditions (i.e., main effects of Entropy Reduction and Specificity). Based on the linking hypothesis outlined above, these results should be interpreted as indexing a facilitation for processing the noun in the HR compared to the LR conditions, and in the OS compared to the MS conditions. This interpretation is in line with the findings in our other measures, mostly those associated with performing the task (ICA, P600, RTs) – but see also the lower N400 for the HR vs. LR conditions. In pattern items, on the other hand, we found a lower fixation rate to the target over the competitor object during the noun in the MS-LR condition (interaction, no main effects). Based on the linking hypothesis from above, this result should be interpreted to index difficulty in processing the noun in the MS-LR condition. This interpretation is, however, incompatible with the results from our other measures. In sum, while in colour items the interpretation of the fixation probability results that is based on the hypothesis linking increased visual attention to facilitated processing matches the findings in our other measures (ICA, P600, RTs, and N400), this is not the case with pattern items, where the interaction found in the MS-LR condition is not reflected in any of our other measures. Other studies have also raised concerns regarding how fixation probabilities are interpreted in visual-world eye-tracking studies (cf. Qing et al., 2018; Sikos et al., 2019a). Our results add to this research, suggesting that cautiousness should generally be exercised with regard to interpreting visual attention as an index of facilitated processing under all circumstances.

6.3 Rational Redundancy

We have so far argued that redundant adjectives may be produced by rational speakers as a means to manage entropy reduction (and, thereby, listener cognitive effort) across their utterances. As we saw in Chapter 2, however, previous research has found mixed results regarding the comprehension of over-specifications. While some studies have found that over-specifications benefit comprehension (Arts et al., 2011a; Brodbeck et al., 2015, as well as Exp. 1-3 in this thesis), others report that they are in fact detrimental (Davies and Katsos, 2013; Engelhardt et al., 2011). Does that mean that over-specifications are not always rational?

6.3.1 Is redundancy always rational?

To the best of our knowledge, Engelhardt et al. (2011) conducted the only other study using ERPs to examine how over-specifications affect processing. Their results show that both colour and size redundant adjectives hindered participants' on-line comprehension. More specifically, they found an N400-like effect (i.e., an effect with a distribution similar to the N400, but peaking later) for the OS compared to the MS conditions, time-locked to the onset of the adjective. This effect indicates difficulty with retrieving the lexical information on the adjective in the OS conditions, and was further supported by reaction time data showing a slowdown in identifying the target object in OS conditions. These results were modulated by the kind of adjectival modifier used in the description: Colour adjectives elicited more positive amplitudes in an early time-window while they also resulted in faster response times compared to size adjectives. So, why were over-specifications problematic in this experiment?

Crucially, in their study the visual context was highly simplified: there were only two objects, differing along two dimensions in each scene (colour and type, or size and type). Additionally, participants were allowed a relatively long preview time (2 sec plus 500 ms with the fixation cross) given how uncomplicated the visual scene was. Thus, by the time participants were presented with the spoken instruction (to look at one of the objects), they could actively predict how each of the two objects in the screen would be referred to were it to be the target. For example, if a visual scene depicted a red star and a blue circle, participants may have already activated the names *star* and *circle*, but not the adjectives *red* and *blue*, in expecting any of the two objects to be the target referent. Therefore, when they eventually encountered the redundant adjective in the OS conditions, this led to a difficulty with retrieving that information. This is why in Engelhardt et al. (2011) the N400 effect is elicited on the adjective, while in our experiments it is observed on the noun; our visual scenes were more complex and participants likely did not have enough time to make predictions about how any of the six objects on them would be referred to. Over-specifications may, therefore, confuse the listeners and result in significant response slowdowns in simple visual contexts.

Despite the disagreement regarding the influence of over-specifications, one thing is common among all referential communication studies: that colour adjectives have a special status. Speakers are more likely to redundantly use colour than other kinds of adjectives (see also Rubio-Fernández, 2016, 2019; Tarenskeen et al., 2015; van Gompel et al., 2019), and listeners are more likely to prefer over-specifications for colour than for other features (see also Sedivy et al., 1999). Davies and Katsos (2013) tested a host of evaluative adjectives, such as *unbroken* or *modern*, among more common size adjectives, such as *tall* and *small*. Participants viewed scenes of four objects and rated the accompanying referring expressions for naturalness. Over-specified expressions were judged to be less natural for describing the target object compared to minimally-specified expressions. Such findings, however, do not suggest that redundancy in general is detrimental to comprehension; they simply suggest that the redundant use of features that are not perceptually salient is dispreferred. In our experiments, we found a preference for colour vs. pattern redundant adjectives by both speakers and listeners, even though pattern, much like colour, is pertinent to the physical appearance of the object it describes and does not evoke a comparison between this object and others. We do, however, wonder whether redundant pattern adjectives would still be dispreferred, in case patterns were similarly salient across objects; e.g., if the dots were equally prominent on all dotted objects (see Appendix A).

In sum, while redundancy may facilitate referential processing in complex visual contexts and for certain types of features, it is not found to be universally useful. Can this evidence be reconciled with the findings from our experiments that over-specifications are beneficial? In what follows we present a unified account of redundancy that views referential communication as a bounded-rational activity.

6.3.2 A Bounded-Rational Account

We propose an account of referential over-specification, according to which, speakers act as bounded-rational agents, aiming for the production of optimal expressions given the conditions in which communication takes place. This means that minimal descriptions are not necessarily deemed optimal. Rather, what the optimal description is depends on the communicative situation: Speakers need to consider the common ground, in order to *minimise joint effort* in establishing reference. In other words, speakers generally aim to strike a balance between their own production effort, and the effort that their addressees will expend to understand the referring expression and act accordingly as determined by an ongoing task. In this sense, *speakers* are likely to produce redundant words in situations that increase the demands for successful interaction, in order to facilitate referential processing for their listeners. Even though it might be more effortful for speakers to produce over-specified compared to minimally-specified utterances (at least in terms of the number of words that need to be articulated), this effort will eventually pay off: Redundancy may facilitate the identification of the target referent, decreasing the likelihood of a misunderstanding that would require the speaker to repeat or revise her utterance. Due to common ground, *listeners* are able to recognise the speakers' intention, and not ascribe other, pragmatic, meaning to the use of redundancy.

In complex visual contexts, for instance, listeners may in fact expect speakers to encode any information that will be useful in their search for the target referent, and not just the information that minimally specifies it. In other words, we propose that aspects of the common ground, such as visual complexity, determine both whether speakers will use redundant adjectives in their expressions, and whether the use of redundancy will hinder or facilitate listeners' comprehension and target identification. We move on to laying out this account in greater detail. We then seek to understand how this account could be extended to include aspects of common ground other than visual complexity that was the focus of this thesis.

In situated communication, the visual scene is part of the speaker's and listener's common ground, and *visual complexity* has been shown to influence the production and comprehension of redundant referring expressions. Previous research has tested the effects of visual clutter, and referential set size on the production of over-specifications (see Clarke et al., 2013; Koolen et al., 2016; Rubio-Fernández, 2016, 2019, among others). Speakers were found to produce long and redundant referring expressions, with highly complex visual scenes, as for instance when they were trying to tell listeners *where's Wally* in the visual scene (Clarke et al., 2013). The likelihood of producing redundant expressions seems to increase with visual complexity (Rubio-Fernández, 2016, 2019). Listeners' preferences match this production behaviour. In complex visual contexts, listeners rely on any information in the referring expression – even if it is redundant – so as to faster restrict the search space and identify the target referent (see Rubio-Fernández, 2020). At the same time, when the visual context is simplified, extra information results in comprehension difficulty for the listeners, as there is no ground in its use (see Davies and Katsos, 2013; Engelhardt et al., 2011). This may explain Sheriff Truman's reaction to Lucy's continuous over-specifications in that 'Twin Peaks' scene at the beginning; with only two telephones in view, it was not clear to the Sheriff why the redundant location information was of any use.

We have argued that complex visual scenes warrant the use of over-specifications, because redundant adjectives may facilitate comprehension processes for the listeners: They provide extra cues to help listeners ground reference, while they also stretch the speakers' message across more linguistic units (time). Such a production strategy is similar to using the optional *that* in cases where a complement clause is not anticipated based on the matrix verb, as maintained by the UID hypothesis (Jaeger, 2010; Levy and Jaeger, 2007) – using *that* speakers avoid a peak in information density at the onset of the complement clause, that could exceed the bounded processing resources of their listeners. In situated communication, as visual complexity increases with the number of referents, so is the uncertainty (entropy) regarding which object will be talked about by the speaker. Words that induce a high reduction in entropy (restriction in the referent set size) are more informative (see Hale, 2003, 2006; Venhuizen et al., 2019), and incur increased processing difficulty (cf. Frank, 2010, 2013). In complex visual scenes speakers may, therefore, encode redundant adjectives so as to avoid peaks in information density that minimal descriptions would otherwise result to. In other words, a preference for over-specifications may be explained by the fact that they optimise for the reduction of referential entropy: By including redundant adjectives, speakers distribute the same bits of information (i.e., cognitive effort) as in minimal descriptions across a longer sequence of words.

Because visual scenes are part of the common ground, visual complexity also affects the listeners' incremental interpretation of referring expressions. That is, visual complexity determines whether after hearing an adjective listeners will consider as referential candidates only objects in contrast pairs (contrastive interpretation), or whether they will consider any object that the adjective selects even if it is singleton (non-contrastive interpretation). Crucially, a contrastive interpretation of the adjective may incur a processing cost on the subsequent noun, if this noun refers to a singleton object (cf. N400 effect for OS in Pattern items), as opposed to a non-contrastive interpretation (cf. no N400 effect for OS in Colour items). The distributional properties of the visual scene, therefore, shape the on-line comprehension of referring expressions. When the demands for successful communication (which are part of the common ground) are relaxed, pragmatic inferences may be generated, leading to a contrastive meaning; when the demands are increased, the interpretation of the adjective may be tuned to a non-contrastive meaning. Thus, the common ground can tune the interpretation of the adjective, affecting comprehension on the noun.

Our experiments directly manipulated the entropy reduction potential of (redundant and necessary) adjectives – i.e., whether they restricted the referential set size to a greater or lesser degree – and have shown that: (a) Adjectives that reduce entropy to a greater extent early on, contributing to a more uniform entropy reduction profile, facilitate the identification of the target referent (see ICA values on the noun, and RTs, with both colour and pattern adjectives). Over-specification produces an effect additive to that (as observed in the P600, ICA and RTs on the noun), even though pattern adjectives were interpreted contrastively (see anticipatory looks to the contrast vs. singleton referent on the adjective, N400 on the noun) unlike colour adjectives. (b) For some speakers, the entropy reduction potential of redundant adjectives affects the production of over-specifications. That is, rational speakers are more likely to include redundant adjectives, when these adjectives reduce entropy at a higher rate, contributing to a more uniform distribution of entropy reduction (i.e., cognitive effort) across their utterances.

All in all, depending on the distributional properties of the visual scene, redundant adjectives may be used in order to contribute to a uniform reduction of entropy (i.e., cognitive effort) across the referring expression. This is depicted schematically in Figure 6.1. In view of the referent set in the middle panel, the speaker can choose from three alternative referring expressions (top panel – speaker) in order to describe the object in the black frame. She can use either (a) a bare noun, (b) a redundant colour adjective, or (c) a redundant pattern adjective. The expression would be minimally-specified under option (a), while it would be over-specified under (b) and (c). Each of these alternatives results in a different entropy reduction profile (see graphs plotting ΔH underneath the speaker's utterances), affecting listeners' identification of the target referent in different ways (see graphs showing listeners' cognitive workload in bottom panel).[1]

The minimal description ('Find the ball'), reduces referential entropy on a single word: 'ball' carries all the information (in bits) for specifying the target referent. This entropy reduction profile results in a peak in information density on the final word (see Fig. 6.1 a – speaker panel), causing increased cognitive effort to the listener at the end of the utterance (see Fig. 6.1 a – listener panel). The two over-specified descriptions ('Find the blue/dotted ball'), on the other hand, distribute the reduction of entropy across two words: the redundant adjective and the head noun. Therefore, the insertion of the redundant adjective results in a more uniform entropy reduction profile compared to that of the minimal description (cf. Fig. 6.1 b and c vs. Fig. 6.1 a in the speaker panel).

The over-specified expressions in Figure 6.1 differ between each other, as well. The redundant use of 'blue', which selects 2 out of 6 referents, results in a more *uniform* entropy reduction profile compared to that of 'dotted', which selects 4 out of 6 referents. As discussed in Experiments 2 and 3, a high reduction of entropy before the final noun (cf. ΔH_{blue} = 1.58 bits in Fig. 6.1 b) results in less residual entropy

[1] The listener plots in Figure 6.1 were produced using the ICA data from Experiment 3.

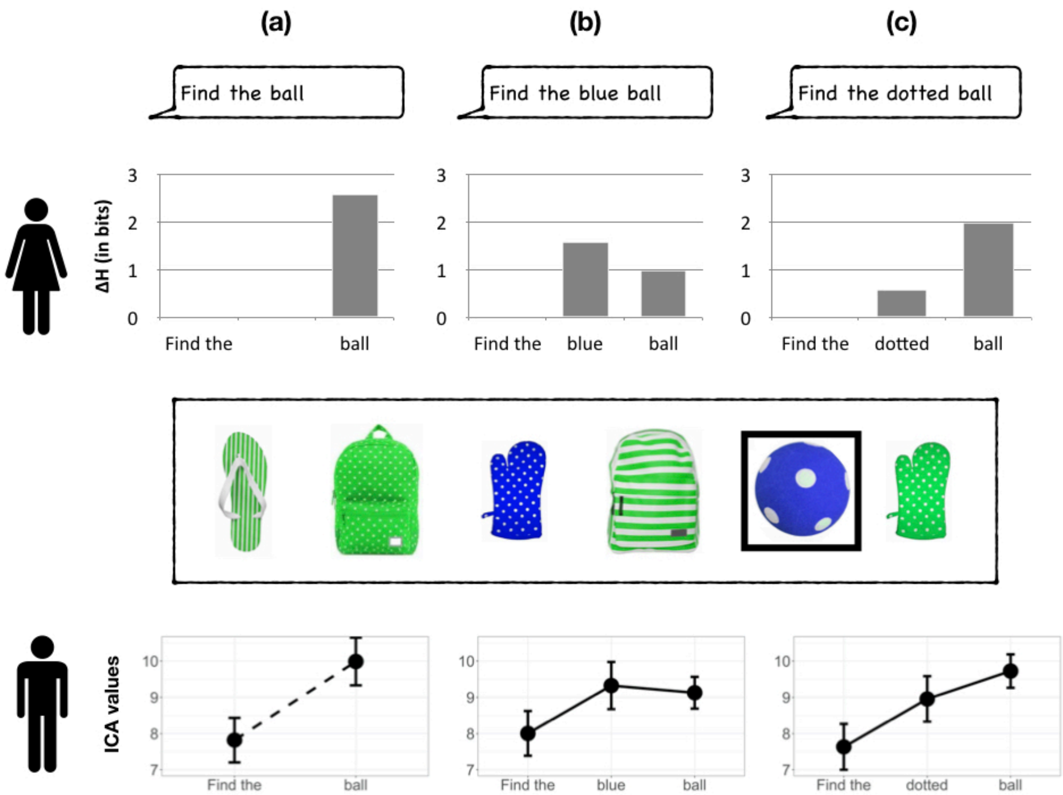


Figure 6.1 Rational redundancy in referential communication.

to be reduced on the noun (cf. ΔH_{ball} = 1 bit in Fig. 6.1 b). By contrast, a small reduction of entropy on the adjective (cf. ΔH_{dotted} = 0.58 bits in Fig. 6.1 c) leaves a higher amount of uncertainty to be reduced on the final noun (cf. ΔH_{ball} = 0.58 bits in Fig. 6.1 c). These different entropy reduction profiles result in differences in the cognitive effort expended by listeners for target identification (see Fig. 6.1 b and c – listener panel). A high reduction of entropy on the adjective results in increased processing effort in this region, but in reduced effort on the subsequent noun (cf. Fig. 6.1 b). By contrast, while a lower reduction of entropy on the adjective is not costly in this region, it is cognitively demanding on the noun, where a larger amount of residual entropy needs to be reduced (cf. Fig. 6.1 b).

The bounded-rational account predicts that, when selecting a referring expression from among (meaning-preserving) alternatives, rational speakers will consider the complexity of the visual context and exploit its distributional properties in order to alleviate their listeners' bounded resources. Rational speakers are, therefore, more likely to encode redundant adjectives in their referring expressions, when these adjectives enable a more uniform distribution of entropy reduction (i.e., cognitive effort) across the utterance. Such a production strategy, which aims at facilitating listeners' target identification effort, also manages speaker effort at the same time: By designing utterances that will benefit their listeners' comprehension and task performance, speakers ensure that their listeners will be able to identify the target referent easily and accurately, and that they will not need to repeat or revise their utterance. Thus, rationally redundant referring expressions allow speakers to successfully establish reference while minimising joint effort. A question is raised, however: What is the tipping point of visual complexity? That is, how many referents are too many, warranting the use of redundant adjectives?

Visual complexity is only one of the aspects in common ground that may influence the production and comprehension of over-specifications, while other aspects may also be at play and influence referential success. Speakers and listeners usually enter a communicative interaction in order to carry out a certain task. The *importance of the shared task* is, therefore, in common ground, and may affect speakers' referential choices. It has been shown, for instance, that speakers who were led to believe that the task they were involved in was highly significant (e.g., a long-distance surgery) produced more over-specifications than speakers who were not given such a cover story (see Arts et al., 2011a; Maes et al., 2004). In other words, speakers encode redundant information, in order to ensure that they provide enough cues to guide their listeners' attention and help them perform the task successfully. Under our account, addressees should also expect their speakers to use more explicit referring expressions in conditions of greater significance, and should, therefore, interpret adjectives literally. Some preliminary evidence in support of this claim comes from a one-shot experiment (Sikos et al., 2019b), where participants completed only one trial guessing which of three objects the speaker aimed to refer to with the use of only one word (e.g., 'blue'). The cover story was that the speaker assisted the participant to disarm a bomb, but as time was running out the speaker could use only one word. The findings show that participants did not use pragmatic reasoning to infer the speaker's intended meaning; a model based on literal meaning and visual salience was found to be better suited to explain participants' responses.

Common ground also includes knowledge of the *partners' conversational skills*. This knowledge may be given (e.g., participants are informed that they will be interacting with a reliable/unreliable partner) or acquired from prior experience with the specific partner (see Brown-Schmidt et al., 2015). For instance, listeners' pragmatic inferencing may depend on the social characteristics of the speaker, e.g., their likeability (Grodner and Sedivy, 2011; Sikos et al., 2019a). Furthermore, listeners are able to adapt expectations regarding the explicitness (or specificity) of a referring expression based on their experience with an over-specifying speaker (Ryskin et al., 2019). This ability may, however, require previous exposure to large amounts of data with the specific partner (cf. Pogue et al., 2016). Similarly, the speaker's prior experience with a particular addressee may affect their upcoming referential choices. If, for instance, a speaker used redundant adjectives to identify singleton targets without receiving negative feedback from the listener (e.g., slowdown in response times or explicit request to use minimal information) in a number of trials, she may infer that redundancy is useful to her listener, and use it again. This hypothesis, however, remains to be investigated.

Environmental conditions that limit the *listener's cognitive capacity* are also part of the common ground. According to our account, conditions that impose limitations on the listeners' cognitive resources, can motivate the use of redundant adjectives by the speakers so as to ease processing and identification of the target referent. It has, in fact, been shown that speakers-passengers describing referents for listener-drivers, who perform a difficult driving task, are more likely to over-specify, when they have first-hand experience with this task themselves (Vogels et al., 2019). Presumably, such cognitively demanding conditions limit the listener's ability to perform complex pragmatic reasoning. When speakers have been in the driver's seat, these cognitive demands are grounded in context. Speakers will, therefore, tune the explicitness of their referring expressions to fit the cognitive state of the listeners, and listeners will utilise any cue in the referring expressions that will ease their performance in the task (driving while searching for the target referent).

Noise in the communication channel is also in common ground. Specifically, speakers' referential choices may be shaped by an effort to compensate for loss of information in their signal, e.g., due to noise in the communicative environment. One relevant example is the 'pilot's alphabet', a communication system used by pilots and air traffic controllers to express combinations of letters and numbers when talking over the radio. In this system, speakers use entire words to represent single letters, e.g., *Charlie* for *C*, thus making sure that, even if part of the signal is corrupted due to noise in the channel, their listener will still be able to recover the original message.

Recently, Degen et al. (2020) put forward an account that explains referential redundancy as a bounded-rational behaviour within the RSA framework (Frank and Goodman, 2012; Goodman and Frank, 2016). This account proposed two modifications in the basic RSA model, in order to permit over-specifications for colour and size. First, the basic RSA model produces only single-word utterances. Degen et al. (2020), therefore, modified the lexicon L, adopting an intersective semantics, thus allowing for the generation of more complex, two-word utterances. As Degen et al. (2020) observe, however, the modification of the lexicon alone does not allow for the production of redundant utterances, and 'the only way for [a] more complex utterance to be chosen with greater probability than [a] simple utterance is if it was the *cheaper* one' (Degen et al., 2020, p.12). Because they consider it unrealistic that more complex utterances are easier than simple utterances (articulating more words leads to increased production effort), they look elsewhere for achieving the same result.

Basic RSA models a pragmatic speaker S, who reasons about a literal listener L_0, and produces an utterance u with probability proportional to the *utility* of that utterance. Speaker's utility depends on both the listener's probability of identifying object o given the specific utterance u (*informativeness*) and the *cost* c of producing that utterance, as given in the equation:

$$U(u,o) = \beta_i ln P_{L_0}(o|u) - \beta_c c(u). \tag{6.1}$$

Listener's informativeness ($P_{L_0}(o|u)$) is computed based on Boolean semantics. That is, the lexicon $L(u,o)$ used by the listener encodes deterministic meanings, taking the value 1 if utterance u is true of object o, and the value 0 in all other cases. In order to account for over-specifications, Degen et al. (2020) relax these semantics, allowing the lexicon to return values between 0 and 1. This *continuous semantics* allows for objects to take on values of varying degrees. That is, instead of capturing objects as unambiguously blue, this altered model allows the degree of 'blue-ness' to vary. In this way, continuous semantics RSA can account for the use of over-specifications.

Our account of over-specification also views redundancy as a bounded-rational behaviour, in that it optimises joint effort in establishing reference, given information that is in common ground. In terms of the Degen et al. model, therefore, our account modifies the *cost* of the utterance, rather than the semantics of the lexicon. As discussed above, the common ground affects the perceived joint effort for successful communication, and influences speakers' choice to utter more explicit (redundant) referring expressions. That is, the cost for producing an over-specified utterance is not an absolute quantity, but is determined relative to the extra effort that speakers (and listeners) would potentially need to expend in case of referential failure. In other words, the common ground determines whether the chance of miscommunication will weigh in with speakers' choice to be redundant. As listeners are exposed to the same conditions, the common ground will also affect their interpretation in a similar way. That is, the common ground will tune the listeners' interpretation of the adjective to a contrastive or literal meaning. There is, therefore, no need to assume an extra, recursive level in the model to account for the cases that listeners employ pragmatic reasoning; this is given 'for free' by the common ground.

To summarise, we suggest that over-specifications may arise as a result of speakers' desire to minimise joint effort for establishing reference while maximising communicative success. Depending on the conditions of the interaction, rational speakers may try to distribute listeners' visual search effort for identifying the target referent across a longer sequence of words. The inclusion of redundant adjectives may serve this goal: The speakers' utterance is stretched over more words, while providing more cues to ensure that the listener will identify the intended referent successfully. Listeners' comprehension is not negatively affected by redundancy, because the conditions that motivate the production of over-specifications are part of the common ground, i.e., they are known to the listener and affect his understanding. More specifically, the common ground may include parameters such as the complexity of the visual context wherein communication takes place, the importance of speakers' and listeners' joint goal, as well as their background knowledge, cognitive state or experience with each other, etc. The common ground may, therefore, tune listeners' interpretation of the adjective to a contrastive or non-contrastive meaning, in turn affecting their processing of over-specifications and their identification of the target referent.

Chapter 7

Conclusions

The present thesis has investigated the influence of the distributional properties of the immediate visual scene on the production and comprehension of redundant expressions referring to objects in the scene. According to Grice's influential theory (Grice, 1975), following a principle of cooperativity rational speakers are expected to avoid unnecessary prolixity. Speakers, however, frequently encode unnecessary adjectives in their descriptions of entities in the visual context, thereby violating this expectation. The current psycholinguistic literature is divided over whether speakers' use of such over-specifications is motivated by production-internal (egocentric) or addressee-oriented (audience-design) concerns, and by extention whether over-specifications hinder addressees' comprehension or not. We have suggested that over-specifications may be produced as a means to ease the addressees' referential processing effort; i.e., the effort associated with the on-line comprehension of the referring expression and the simultaneous visual search for grounding reference in context.

Recently, measures of informativity, such as the surprisal (expectancy) or reduction of entropy (uncertainty), have been adopted in linguistic research in order to quantify cognitive load in language processing (Hale, 2003, 2006; Levy, 2008). These information-theoretic accounts argue that the surprisal of a word (Hale, 2001; Levy, 2008; Smith and Levy, 2013), or the reduction of entropy induced by a word (Frank, 2013; Hale, 2003, 2006), measured in bits of information, are proportional to the cognitive effort associated with processing that word. Similarly, information-theoretic accounts of language production suggest that cooperative speakers choose encodings that distribute processing effort across more linguistic units (Jaeger, 2010; Levy and Jaeger, 2007).

We extended the use of such information-theoretic metrics into visually-situated communication, and propose that rational and cooperative speakers may encode redundant adjectives, when they believe that these adjectives can ease listeners' cognitive effort and facilitate referential success: By including redundant adjectives, speakers distribute the reduction of referential entropy – and thereby cognitive effort – across more words. Because the conditions that determine this choice for the speaker are part of the common ground they share with the listener, listener's expectations regarding the use of redundant information are shifted. The common ground, therefore, motivates the use of redundant adjectives and influences their on-line interpretation.

We have explored the predictions of this bounded-rational account in a series of comprehension and production experiments. Since there is as of yet no conclusive evidence regarding the comprehension of over-specifications, we began by investigating exactly this: how over-specifications affect listeners' on-line comprehension and visual search for the target referent. We then moved on to examining how the distributional properties of the immediate visual scene affect comprehension processes, and finally whether these properties, as well as listeners' preferences are taken into account by speakers when planning their utterances. Taken together, our results reveal that referential redundancy can be rational, in that, even if on-line comprehension is hindered, it actually facilitates listeners' task performance, and that cooperative speakers may utilise it in order to help listeners navigate the visual scene and minimise joint effort for establishing reference.

In Experiment 1, we found evidence that, in the presence of complex visual scenes, listeners' on-line comprehension is actually facilitated by the use of redundant adjectives: noun processing was benefited after a redundant adjective. Because the target referent was singleton only in the visual scenes of the over-specified condition, it is possible that the observed benefit was a result of the visually-determined surprisal on the noun. This issue was addressed in Experiments 2 and 3, by having on every display two objects matching the type of the target referent: one in a contrast pair (contrast object) and one singleton.

Experiments 2 and 3, moreover, examined the influence of referential entropy reduction on comprehension processes. Crucially, these experiments yielded different effects with colour and with pattern adjectives. Anticipatory eye movements to the contrast object, as well as an N400 effect for the over-specified noun, indicated that pattern adjectives were interpreted contrastively, in line with the Gricean account. By contrast, the absence of such results with colour adjectives suggests that they were interpreted literally. Interestingly, however, the lower ICA and enhanced P600 found on the noun following both colour and pattern redundant adjectives indicates that participants' visual search for the target referent and performance in the task were facilitated by redundancy. As predicted by the Entropy Reduction Hypothesis (Hale, 2003, 2006), a higher reduction of entropy on the adjective resulted in increased cognitive effort in that region, but lower effort on the following noun. This effect suggests that restricting the visual search space early on facilitates the identification of the target referent.

As Experiments 2 and 3 tested the same set of stimuli using different methodologies (ERP/ICA and eye-tracking/ICA, respectively), their results are important from a methodological perspective, as well. Specifically, these findings tap into two aspects of visually-situated communication: linguistic and situational. The linguistic aspects are related to the *on-line understanding* of the referring expression, and were indexed by the anticipatory eye movements observed on the adjective, and the N400 effect on the noun. The situational aspects are associated with the (linguistically-guided) visual search for grounding reference in context and *performing the task*, and were indexed by the ICA on the adjective and the noun, as well as the P600 on the noun and the overall RTs.

Taken together, Experiments 1-3 exhibit several important insights: (a) that information-theoretic measures, such as surprisal and entropy reduction, can be extended into visual-world studies to quantify listeners' cognitive workload; (b) that when visual complexity is high, redundant adjectives may facilitate the on-line comprehension of the referring expression, by guiding listeners' visual search for the referent; (c) that the advantage observed for referential redundancy largely depends on the kind of adjective used (colour adjectives are preferred).

Our next two experiments turned to the production of over-specifications, and investigated whether speakers are sensitive to the distributional properties of the visual scene, and choose to over-specify for adjectives with a high entropy-reduction potential; i.e., whether the production of over-specifications is motivated by an incentive to manage the rate of entropy reduction for the addressee. Both experiments found a great degree of diversity among speakers' production choices, and a general preference for colour over pattern over-specifications. In both experiments, we identified three groups of speakers based on their use of over-specification. Interestingly, our results are in line with both egocentric and audience-design accounts of reference production. Rational speakers over-specified more for adjectives that reduced entropy at a high rate when no adjective was necessary (Experiment 4) or for adjectives that reduced entropy at a higher rate than the necessary adjective (Experiment 5), thus effecting a more uniform reduction of entropy (i.e., cognitive load) across their utterances.

All in all, adjectives need not evoke pragmatic inferencing, i.e., imply a comparison between two same-type objects. Their functionality depends on aspects of the common ground such as visual complexity, as discussed above, but also on factors that are independent of the communicative situation, such as the language used or the register. For instance, Dye et al. (2018) have shown that adjectives in languages like English that have evolved with a probabilistic determiner system (i.e., not encoding information such as gender) may frequently be used redundantly, because they help manage uncertainty about the rest of the sentence, as do determiners in languages like German. Additionally, depending on the register adjectives may serve different functionalities. In literature, for instance, adjectives may be purely ornamental. No one would argue that a sentence such as 'The man lived alone in the old house' in a novel implies the existence of a *new* house. There is, furthermore, evidence from child-directed speech that redundant adjectives not only facilitate reference establishment, as discussed throughout this thesis, but also enhance the child's language learning (Arunachalam, 2016; Davies et al., 2020). Most importantly, the *rational* use of redundancy seems to be a skill like any other that is learned at a young age (Deutsch and Pechmann, 1982) and declines as speakers get older (cf. Saryazdi et al., 2019, where older adults over-specified with uncommon adjectives).

In sum, in visually-situated communication, too, listeners are sensitive to the distribution of information across the signal. Rational speakers are, therefore, likely to utilise redundancy in order to distribute information more evenly across their utterances, and maximise efficiency while minimising joint effort.

Chapter 8

Zusammenfassung

Referentielle Redundanz stellt ein Rätsel für Grices pragmatische Kommunikationsbeschreibungen dar, nach denen Sprecher, die einen Zielreferenten in einer visuellen Szene zu spezifizieren beschreiben, ausschließlich die notwendigen Informationen produzieren sollten. Da die Adressaten erwarten, dass sich die Sprecher an solche Gesprächsnormen halten, sollte ihr Verständnis durch redundante Informationen gestört werden. Mit anderen Worten: Rationale Sprecher sollten immer minimale Beschreibungen wählen, um einen beabsichtigten Referenten zu spezifizieren, und rationale Zuhörer sollten immer erwarten, dass die Informationen, die sie erhalten, für diese Aufgabe notwendig sind. Empirische Studien haben jedoch immer wieder gezeigt, dass Sprecher in ihren referierenden Äußerungen tatsächlich recht häufig unnötige Informationen kodieren, was die Frage aufwirft: *warum die Sprecher überspezifizieren?*

Zur Beantwortung dieser Frage gibt es im Allgemeinen zwei Ansätze. Nach der *Egozentrische Darstellung* zielen die Sprecher vor allem darauf ab, den Produktionsaufwand zu reduzieren, was dazu führen kann, dass sie mit dem Sprechen beginnen, bevor sie die visuelle Szene erschöpfend nach möglichen mit dem Ziel konkurrierenden Objekten durchsucht haben. Daher könnten Äußerungen Informationen enthalten, die sich am Ende als unnötig erweisen. Nach der *Publikumsdesign Darstellung* hingegen zielen die Überspezifikationen der Sprecher darauf ab, ihren Adressaten die referenzielle Verarbeitung zu erleichtern. Die Entscheidung zwischen diesen Ansätzen hängt daher stark davon ab, wie die Adressaten auf die Überspezifikationen reagieren. Das in der Literatur beschriebene Verständnis von Überspezifikationen ist jedoch uneinheitlich. Einige Studien stellten fest, dass Überspezifikationen das Verständnis behindern, während andere zeigen, dass sie es im Gegenteil erleichtern könnten.

In dieser Arbeit wurde untersucht, ob eine begrenzt-rationale Darstellung der Kommunikation besser geeignet ist, um die Verwendung von Redundanz zu erklären und die scheinbar unvereinbaren empirischen Belege miteinander in Einklang zu bringen. Wir schlagen vor, dass rationale Sprecher danach streben sollten, die *gemeinsame Anstrengung* bei der Produktion von Referenzen zu minimieren, indem sie redundante Beschreibungen erstellen, wenn die zusätzlichen Informationen dazu beitragen, die Unsicherheit über den Zielreferenten effizienter (d.h. schnell, einfach und zuverlässig) für ihre Zuhörer zu reduzieren. Mit anderen Worten, minimale Beschreibungen sind nicht immer optimal; je nach Faktoren wie der Komplexität der visuellen Szene, der Wichtigkeit der verbundenen Aufgabe usw. können redundante Informationen der referentiellen Kommunikation zugute kommen.

Neuere Kommunikationsansätze, die auf der Informationstheorie (Shannon, 1948) basieren, haben gezeigt, dass wir die Informativität eines Wortes als die Erwartung dieses Wortes in seinem Kontext (Überraschung; siehe Hale, 2001; Levy, 2008) messen können, oder als das Ausmaß der Verringerung der Unsicherheit über den Rest des Satzes (Entropie), die dieses Wort (siehe Hale, 2003, 2006) mit sich bringt. Diese Informationswerte wurden mit dem Verarbeitungsaufwand in Verbindung gebracht: Je höher die Information in Bits eines bestimmten Worts (höhere Überraschung oder Verringerung der Entropie), desto mehr kognitive Anstrengung ist für die Verarbeitung dieses Wortes (vgl. Frank, 2013; Smith and Levy, 2013) erforderlich. Die Uniform Information Density (UID)-Hypothese sagt demnach voraus, dass die Produktionsentscheidungen der Sprecher durch die Absicht einer gleichmäßigen Verteilung von Informationen (und damit der kognitiven Anstrengung) über mehr sprachliche Einheiten (Jaeger, 2010; Levy and Jaeger, 2007) bestimmt werden können. Das heißt, wenn sie mit zwei bedeutungserhaltenden syntaktischen Alternativen zum Ausdrücken einer Botschaft konfrontiert sind, wählen die Sprecher eher diejenige, die die Information gleichmäßiger über die Äußerung verteilt, um Spitzen in der kognitiven Belastung zu vermeiden, die die begrenzten Verarbeitungsressourcen ihrer Zuhörer übersteigen könnten. Beispielsweise können die Sprecher bei der Planung eines Komplementärsatzes die Unterkategorisierungspräferenzen eines Verbs berücksichtigen: Falls ein Komplementärsatz (z.B. 'wir waren absolut verrückt') nicht mit hoher Wahrscheinlichkeit nach dem Hauptverb (z.B. 'bestätigt') folgt, können die Sprecher sich dafür entscheiden, das optionale *that* (dass) einzufügen.

Andernfalls wäre das erste Wort im Komplementärsatz (z.B. 'wir') zwischen dem Beginn des Satzes und seinem semantischen Inhalt mehrdeutig und somit sehr informativ. Die Sprecher können daher Kodierungen wählen, die den Informationsgehalt über mehr sprachliche Einheiten (und Zeit) verteilen und so vermeiden, dass informationsdichte Spitzen in ihren Äußerungen zu Verarbeitungsschwierigkeiten für die Zuhörer führen.

Bestehende Ansätze messen die Informativität, anhand des sprachlichen Kontexts. Wir haben diese auf visuell-situative Kommunikation erweitert, bei der die Informativität eines Wortes auf der Grundlage seiner Wahrscheinlichkeit, sowohl in der aktuellen Äußerung als auch in der jeweiligen visuellen Szene bestimmt wird. Dazu führten wir den Begriff der *Referenzentropie* als ein Maß für die Ungewissheit ein, welches Objekt in einer visuellen Szene in der referierenden Äußerung erwähnt wird. Zum Beispiel ist die referentielle Entropie bei '*blauen*' in 'Finde den blauen. . . ' in einer visuellen Szene mit zwei blauen Objekten niedriger als in einer Szene mit vier blauen Objekten, obwohl in beiden Fällen die Äußerung genau gleich ist. Das heißt, '*blauen*' reduziert die referentielle Entropie in unterschiedlichem Maße, je nach der Verfügbarkeit von blauen Objekten im visuellen Kontext. Durch Erweiterung der Entropie-Reduktionshypothese (Hale, 2003, 2006) auf die visuelle Welt (d.h. Messung der Reduktion der Entropie über potentielle Referenten in einer visuellen Szene anstelle möglicher Satzfortsetzungen) sollte der kognitive Aufwand für die Verarbeitung eines Wortes proportional zum Grad der durch dieses Wort induzierten referentiellen Entropie-Reduktion in Informationsbits sein. Um erfolgreich einen Bezug herzustellen, muss die referentielle Entropie natürlich auf null reduziert werden, und die Sprecher können ihre Äußerungen so formulieren, dass die Rate, mit der die anfängliche Entropie minimiert wird, moduliert wird.

Wir schlagen daher eine begrenzt-rationale Darstellung der referentiellen Redundanz vor, nach der jedes Wort in einer Äußerung, auch wenn es redundant ist, dem Verständnis förderlich ist, solange es zur Verringerung der referentiellen Entropie beiträgt. Im Rahmen dieses Ansatzes können die Sprecher überspezifikationen nutzen, um den Aufwand für die Verarbeitung des referentiellen Ausdrucks und die Identifizierung des Zielreferenten gleichmäßiger auf die Aussage zu verteilen. Zur Beurteilung dieses Berichts haben wir drei Forschungsfragen entwickelt, die jeweils auf der Antwort auf die vorhergehende Frage aufbauen, und eine Reihe von Verständnis- und Produktionsexperimenten entworfen, die sich mit jeder dieser Fragen befassen. Unsere Forschungsfragen waren:

FF-1 Ist Überspezifizierung dem Verständnis abträglich oder förderlich?

FF-2 Wird das Verständnis von Referenzausdrücken dadurch beeinflusst, wie effizient sie die Reduktion der referentiellen Entropie bewältigen?

FF-3 Sind Sprecher empfindlich für die Verteilungseigenschaften der visuellen Szene, und wenn ja, verwenden sie Überspezifikationen als Mittel zur Modulation der referentiellen Entropie für ihre Zuhörer?

Ein erstes ereigniskorrelierte Potenziale (EKP) Experiment befasste sich mit FF-1. Den Teilnehmern wurden komplexe visuelle Szenen präsentiert, in denen sechs Objekte dargestellt wurden, die sich in drei Dimensionen (Farbe, Muster und Typ) unterschieden, gepaart mit einer Audioanweisung zur Lokalisierung eines der Objekte in der Szene. Wir manipulierten die Spezifität des referentiellen Ausdrucks: Während die Audioanweisung (z.B. 'Finde die gelbe Schüssel') unter verschiedenen Bedingungen konstant gehalten wurde, unterschieden sich die visuellen Szenen, wobei die Äußerung minimal spezifiziert (Minimally Specified; MS), über-spezifiziert (Over-specified; OS) oder unterspezifiziert (Under-specified; US) wiedergegeben wurde. Im MS-Zustand stellte die visuelle Szene zwei Objekte desselben Typs (z.B. zwei Schalen) dar, die sich in einem Merkmal (Farbe oder Muster) unterschieden, und die Äußerung trug daher die minimale Information zur Identifizierung des Zielreferenten. In der OS-Bedingung war nur ein Objekt des genannten Typs in der visuellen Szene vorhanden (z.B. eine Schale), und daher war das Adjektiv im referierenden Ausdruck redundant. Sowohl in der MS-Bedingung als auch in der OS-Bedingung konnte der Bezug jedoch erfolgreich hergestellt werden; es war eindeutig, welches Objekt der Zielreferent war. Die US-Bedingung hingegen führte zum Scheitern der Referenz: Obwohl die Äußerung den Typ des Zielreferenten identifizierte, half sie den Zuhörern nicht, zwischen den beiden in der Szene vorhandenen Objekten gleichen Typs (z.B. zwei gelbe Schalen) zu unterscheiden. Um festzustellen, ob ein referentielles Versagen aufgrund von Unterspezifikation ähnlich ist wie ein *explizites* referentielles Versagen aufgrund einer Diskrepanz zwischen der visuellen und der sprachlichen Eingabe, wurde auch eine Mismatch (MM)-Bedingung verwendet. In dieser Bedingung deutete das Adjektiv auf einen einzelnen Referenten him (d.h. es gab nur ein gelbes Objekt in der visuellen Szene), welches jedoch nicht mit dem genannten Substantiv übereinstimmte.

Unsere Ergebnisse lieferten zwei wichtige Erkenntnisse dar. Erstens unterschieden unsere Ergebnisse zwei qualitativ unterschiedliche Prozesse, die mit dem referentiellen Versagen verbunden sind: referentielles Versagen aufgrund des Mangels an

Information und referentielles Versagen aufgrund des Missverhältnisses zwischen sprachlicher und visueller Information. Insbesondere ergab die US-Bedingung im Vergleich zur MS-Bedingung eine lang anhaltende Positivität, die etwa 400 ms nach dem Auftreten des Adjektivs begann und während des gesamten Zeitfensters des Substantivs anhielt. Die MM-Bedingung hingegen verursachte einen großen N400-Effekt im Substantiv-Zeitfenster (siehe unten).

Zweitens - als zentrales Anliegen dieser Arbeit - wurden in komplexen visuellen Kontexten Überspezifikationen festgestellt, die das Verständnis begünstigen. Insbesondere beobachteten wir einen abgestuften N400-Effekt für die MS-, OS- und MM-Bedingungen im Substantivbereich, wobei die OS-Bedingung die am stärksten reduzierte N400-Amplitude und die MM-Bedingung die höchste N400-Amplitude auslöste. Da die N400-Komponente mit der Worterwartung in Verbindung gebracht wurde, interpretierten wir diesen abgestuften N400-Effekt als Beweis für die visuell bedingte Überraschung auf dem Substantiv. Das heißt, während in der MM-Bedingung das Substantiv (z.B. 'Gießkanne') aufgrund des vorhergehenden Adjektivs (z.B. 'gelb') unerwartet war, war in der OS-Bedingung der durch das Substantiv beschriebene Referent (z.B. Schüssel) der einzige Referent in der Szene, der mit dem Adjektiv übereinstimmte. Daher war er im Vergleich zur MS-Bedingung, bei der das Adjektiv zwei Referenten auswählte, in hohem Maße vorhersehbar. Dieser Befund steht im Gegensatz zu einer streng rationalen (Gricean) Darstellung der Kommunikation, entspricht aber unserer begrenzt-rationalen Darstellung.

In den Experimenten 2 und 3 wurde untersucht, ob die Rate, mit der die referentielle Entropie über einen referierenden Ausdruck reduziert wird, auch die Verarbeitung beeinflusst und ob die Entropie-Reduktion und die Spezifität additive Effekte haben (FF-2). Das heißt, in diesen Experimenten wurde untersucht, ob Ausdrücke, die die referentielle Entropie effizienter reduzieren (d.h. die Reduktion der referentiellen Entropie gleichmäßiger über die Äußerung verteilen), die Verarbeitung begünstigen, und ob Effekte der Spezifität über die Effekte der Entropie-Reduktion hinaus beobachtet werden können. Es ist jedoch möglich, dass die bei den Überspezifikationen im Vergleich zu den Minimalbeschreibungen in Experiment 1 beobachtete Erleichterung ein Nebenprodukt der Anzeigestruktur im OS-Zustand dieses Experiments war, bei dem nur ein Objekt mit dem Adjektiv übereinstimmte und dieses Objekt ein einzigartiges Objekt war. Das heißt, die visuellen Szenen in der OS-Bedingung unterstützten nur eine nicht-kontrastierende Interpretation des Adjektivs. Dieses Problem wurde auch in den Experimenten 2 und 3 behandelt, indem visuelle Szenen verwendet wurden, die sowohl eine kontrastive als auch eine nicht-kontrastive Lesart des Adjektivs unterstützten; d.h., sowohl Referenzpunkte in Kontrastpaaren als auch Singleton-Referenzen passten zum Adjektiv und waren gleichzeitig in den Szenen verfügbar. Wir erstellten eine Reihe von Stimuli, die alle in einem ERP-Experiment (Exp. 2) und zur Hälfte in einem Eye-Tracking-Experiment (Exp. 3) verwendet wurden. In beiden Experimenten maßen wir außerdem den Index der kognitiven Aktivität (Index of Cognitive Activity; ICA), der ein Pupillenmaß ist und sowohl in einem ERP-Experiment als auch in einem Eye-Tracking-Experiment als Index der gesamten kognitiven Anstrengung aufgezeichnet werden kann. Wir erwarteten, dass die Ergebnisse des ICA in den beiden Experimenten ähnlich sein würden, so dass es möglich wäre, das ERP mit den Eye-Tracking-Ergebnissen zu korrelieren. Vier visuelle Darstellungen wurden mit einer gesprochenen Anweisung gepaart. Die Anweisung war minimal spezifiziert (MS), wenn sie ein Objekt innerhalb eines Kontrastpaares identifizierte, und wurde überspezifiziert (OS), wenn sie ein einzigartiges Objekt identifizierte. Darüber hinaus wurde die referentielle Entropie mit einer höheren Rate reduziert (High Reduction; HR), wenn das pränominale Adjektiv zwei potenzielle Referenten auswählte, während sie mit einer niedrigeren Rate reduziert wurde (Low Reduction; LR), wenn das Adjektiv vier Referenten auswählte. Auf diese Weise wurden vier experimentelle Bedingungen erzeugt: MS-HR, MS-LR, OS-HR und OS-LR.

Die Ergebnisse dieser Experimente zeichnen ein komplexes Bild. Erstens fanden wir Hinweise darauf, dass eine Überspezifizierung das Verständnis auf unterschiedliche Weise beeinträchtigt, je nach Art des überflüssigen Adjektivs, das in dem Ausdruck verwendet wird. Redundante Musteradjektive wurden inkrementell kontrastiv interpretiert (mehr vorausschauende Augenbewegungen zum Kontrast vs. einzigartige Objekte während des Adjektivs), was die Verarbeitung des folgenden Substantivs erschwerte (N400-Effekt für die OS- vs. MS-Bedingungen). Andererseits wurde bei Farbadjektiven weder eine kontrastive Lesart gefunden (kein Unterschied in den antizipativen Augenbewegungen zum Kontrast vs. einzigartige Objekte während des Adjektivs), noch wurde die Verarbeitung des Substantivs negativ beeinflusst (kein Unterschied in N400 zwischen den OS- und den MS-Bedingungen).

Gleichzeitig wurden es evident, dass beide Arten von redundanten Adjektiven (Farbe und Muster) die *visuelle Suche* der Zuhörer zur Identifizierung des Zielreferenten und zur Durchführung der Aufgabe erleichterten. Das heißt, selbst wenn das Online-Verständnis durch die Redundanz behindert wurde, bot das redundante

Adjektiv zusätzliche visuelle Hinweise, die es den Zuhörern ermöglichten, die Menge der potentiellen Referenten einzuschränken und die Ausführung der Aufgabe zu erleichtern (Angabe, auf welcher Seite des Bildschirms der Zielreferent erschien). Diese Interpretation wird durch ERP-Daten (P600-Effekt für die OS-Bedingungen), Eye-Tracking (mehr Fixierungen auf das Zielobjekt im Vergleich zu Konkurrenzobjekten bei OS- vs. MS-Bedingungen) und Pupillendaten (niedrigere ICA-Werte bei OS- vs. MS-Bedingungen während des Substantivs) in der Substantivregion sowie durch Verhaltensmerkmale (schnellere Reaktionszeiten bei OS- vs. MS-Bedingungen) unterstützt. Diese Effekte wurden sowohl bei farb- als auch musterredundanten Adjektiven beobachtet, waren aber bei den Farbadjektiven stärker.

Darüber hinaus zeigten unsere Ergebnisse, dass die Rate, mit der die referentielle Entropie während der Äußerung reduziert wird, ebenfalls Verständnisprozesse beeinflusst. In Einklang mit der Entropie-Reduktions-Hypothese (Hale, 2003, 2006) führte eine hohe Rate der Entropie-Reduktion beim (redundanten oder notwendigen) Adjektiv (d.h. ein Adjektiv, das 2 von 6 Referenten auswählt, im Vergleich zu 4 von 6 Referenten) zu einem größeren kognitiven Aufwand in dieser Region, aber zu einem geringeren Aufwand beim folgenden Substantiv. Interessanterweise war dies ein Effekt, der nur bei den ICA-Maßen von Experiment 3 beobachtet wurde (d.h. nur bei der Blickverfolgung und nicht beim EKP-Experiment). Da der einzige Unterschied zwischen den beiden Experimenten darin bestand, dass die Teilnehmer im EKP-Experiment die visuelle Szene nicht frei abtasten konnten, während sie es im Eye-Tracking-Experiment konnten, glauben wir, dass die kognitive Anstrengung, die durch ICA indiziert wird, von der Schwierigkeit der visuell begründeten Identifizierung des Zielreferenten herrührt. Mit anderen Worten, der ICA erfasst die kognitive Anstrengung des Scannens der visuellen Szene, um den Satz potenzieller Referenten einzugrenzen und den Standort des Ziels zu identifizieren. Diese Schwierigkeit war beim Adjektiv höher, aber beim Substantiv für die HR geringer als bei den LR-Bedingungen (es sei daran erinnert, dass die Auswirkungen der Entropie-Reduktion auf die ICA sowohl bei Farb- als auch bei Musterelementen ähnlich waren). Zusätzlich zu den Auswirkungen auf den ICA fanden wir auch, dass eine hohe Entropie-Reduktion auf dem (redundanten oder notwendigen) Adjektiv die Verarbeitung auf dem nachfolgenden Substantiv erleichterte (reduzierter N400-Effekt für HR vs. LR) und die Ziel-Identifikationszeiten der Teilnehmer verbesserte (schnellere Reaktionszeiten in HR vs. LR). Eine Wechselwirkung zwischen Spezifität und Entropie-Reduktion wurde ebenfalls beobachtet, jedoch nur bei einer unserer Messungen (EKP) und nur mit Farbadjektiven. D.h. es wurde festgestellt, dass redundante Farbadjektive die Leistung der Teilnehmer besonders dann verbessern, wenn sie zu einer höheren Reduktion der referentiellen Entropie führen (P600-Effekt für die OS-HR-Bedingung auf dem Substantiv). Schließlich konnten wir mit Hilfe der ICA den kognitiven Aufwand für die Verarbeitung von überspezifizierten Äußerungen (z.B. 'Finde den blauen Ball' in einer visuellen Szene mit nur einem Ball; OS-HR-Bedingung) direkt mit dem für ihre minimal spezifizierten Gegensätze (z.B. 'Finde den Ball' in derselben Szene; MS-Fillers) vergleichen.Dieser Vergleich zeigte, dass der kognitive Aufwand der Zuhörer zur Identifizierung des Zielreferenten bei Substantiven, die redundanten Adjektiven folgen (OS-HR-Bedingung), geringer war als bei alleinstehenden Substantiven (MS-Fillers).

Zusammenfassend haben unsere Verständnisexperimente gezeigt, dass je nach visuellem Kontext Überspezifikationen die Verarbeitung erleichtern können. In visuellen Kontexten, die sowohl eine kontrastive als auch eine nicht-kontrastive Interpretation des Adjektivs unterstützen (d.h., wenn mindestens zwei Objekte mit dem Adjektiv übereinstimmen: eines in einem Kontrastpaar und ein Unikat), wurde das Online-Verständnis durch redundante Musteradjektive, nicht aber durch redundante Farbadjektive (antizipative Fixierungen auf Unikat- vs. Kontrastobjekte auf dem Adjektiv und N400 auf dem Substantiv) behindert. Dieser Befund stimmt sowohl mit dem Gricean (Musteradjektive) als auch mit der Beschreibung der begrenzt-rationalen Darstellung (Farbadjektive) überein. Die visuelle Suche der Zuhörer nach dem Zielreferenten wurde jedoch *unabhängig* von der Art des verwendeten Adjektivs durch eine referentielle Redundanz begünstigt (P600 und ICA auf dem Substantiv, Gesamt-Reaktionszeiten), was die begrenzt-rationale Darstellung unterstützt. In jedem Fall wurde festgestellt, dass die Rate, mit der die referentielle Entropie während der Äußerung reduziert wird, das Verständnis weiter beeinflusst. Eine hohe Rate der Entropie-Reduktion beim Adjektiv führte zu Schwierigkeiten in diesem Bereich, erleichterte aber die Verarbeitung und visuelle Erdung (grounding) des referentiellen Ausdrucks auf dem folgenden Substantiv (reduziertes N400, ICA, Reaktionszeiten).Schließlich begünstigten redundante Farbadjektive, die eine hohe Rate der Entropie-Reduktion induzierten, besonders die Leistung der Teilnehmer (P600 für die OS-HR-Bedingung auf dem Substantiv).

Nachdem wir ein Verständnis dafür entwickelt hatten, wie Überspezifikationen online verarbeitet werden und wie die referentielle Entropie-Reduktion die Verarbeitung weiter beeinflusst, gingen wir zum Produktionsende der referentiellen Kommunikation über und prüften, ob solche Verständnispräferenzen von den Referenten berücksichtigt werden. In zwei Produktionsexperimenten untersuchten wir die Verwendung von Überspezifikationen durch die Sprecher in visuellen Kontexten, in denen kein Adjektiv zur Identifizierung des Zielreferenten erforderlich war (Exp. 4), und in Kontexten, in denen nur ein Adjektiv erforderlich war (Exp. 5). In beiden Experimenten manipulierten wir das Entropie-Reduktionspotential eines redundanten Adjektivs: ob ein redundantes Farbadjektiv die referentielle Entropie im Vergleich zu einem redundanten Musteradjektiv mehr oder weniger reduzieren würde oder ob beide die Entropie in gleichem Maße reduzieren würden. Daraus ergaben sich drei experimentelle Bedingungen: Farbe-reduzierend, Muster-reduzierend und Gleichwertig-reduzierend. In Experiment 5 wurden diese Bedingungen mit der notwendigen Adjektivmanipulation gekreuzt, d.h., ob eine Farbe oder ein Musteradjektiv zur Identifizierung des Zielreferenten notwendig war.

Wir erwarteten im Allgemeinen, dass unsere Ergebnisse entweder mit den Egozentrischen oder den Publikumsdesign Ansätzen übereinstimmen würden. Nach der egozentrischen Sichtweise spezifizieren die Referenten zu viel, um produktionsinterne Prozesse zu erleichtern. Daher wird erwartet, dass die Auswahl der Referenten nicht durch die Verteilung der Objekte in der visuellen Szene beeinflusst wird und dass ihre OS-Rate unter den verschiedenen Bedingungen nicht variiert. Sowohl die Gricean als auch die begrenzt-rationale Darstellung fallen unter die Sichtweise des Publikumsdesigns. Auf der Grundlage des Gricean-Berichts sollten die Sprecher nur minimale Beschreibungen erstellen, da ihre Zuhörer dies erwarten, und eine zu hohe Spezifizierung würde daher das Verständnis stören. Auf der Grundlage der begrenzt-rationalen Darstellung wird hingegen erwartet, dass Sprecher redundante Adjektive verwenden, wenn sie ein allgemein hohes Entropie-Reduktionspotential haben (Exp. 4) oder wenn sie mehr Entropie-Reduktion als das notwendige Adjektiv aufweisen (Exp. 5). Darüber hinaus erwarteten wir aufgrund der Ergebnisse unserer Verständnisstudien, dass die Farb-OS-Rate insgesamt höher sein sollte als die Muster-OS-Rate, da sich herausstellte, dass die Zuhörer eine Präferenz für farb- gegenüber musterredundanten Adjektiven haben.

Die Ergebnisse liefern zwei wichtige Erkenntnisse hinsichtlich der Produktion von Überspezifikationen. Erstens, dass die Produktionsentscheidungen der Sprecher sehr unterschiedlich sind; in beiden Experimenten haben wir Gruppen von Sprechern identifiziert, die sich für unterschiedliche Produktionsstrategien entschieden haben. Zweitens fanden wir heraus, dass zumindest bei einigen der Sprecher die Wahl der

Referenten von dem Bestreben geleitet sein kann, die Entropie in ihren Äußerungen effizient zu reduzieren, was die begrenzt-rationale Darstellung unterstützt. Genauer gesagt konnten wir in beiden Experimenten die Teilnehmer auf der Grundlage ihrer Produktionsentscheidungen in Gruppen einteilen. In Experiment 4 produzierte eine erste Gruppe von Sprechern über 80% der Zeit unter allen Bedingungen Überspezifikationen mit entweder Farbe, Muster oder beiden Adjektiven, was der egozentrischen Sichtweise entspricht. Eine zweite Gruppe von Sprechern produzierte fast ausschließlich (mehr als 90%) minimale Beschreibungen unter allen Bedingungen. Diese Sprecher wählten eine Gricean-Produktionsstrategie. Der Rest der Sprecher wurde in Gruppen auch wiederum gruppiert, und es wurde festgestellt, dass sie häufiger überspezifizierten, wenn das überflüssige Adjektiv ein hohes Entropie-Reduktionspotential hatte, was der begrenzt-rationalen Darstellung entsprach.

In ähnlicher Weise identifizierten wir in Experiment 5 eine erste Gruppe von Sprechern, die unabhängig von der Bedingung mehr als 80% der Zeit redundante Adjektive produzierten, wie es die egozentrische Darstellung voraussagt. Eine zweite Gruppe wurde von Sprechern gebildet, die in mehr als 80% der Fälle überspezifizierte Farbadjektive protuzierten. Dies ist eine Strategie, die sowohl mit den egozentrischen als auch mit den begrenzt-rationalen Darstellungen übereinstimmt: Farbe ist ein hervorstechendes Merkmal und zieht die Aufmerksamkeit der Sprecher auf sich, die dadurch in der Lage sind, ihre Äußerungen damit zu beginnen und möglicherweise zusätzliche Zeit zu gewinnen, um festzustellen, welche anderen Merkmale - wenn überhaupt - sie erwähnen müssen. Gleichzeitig kann Farbe aufgrund ihrer Auffälligkeit die visuelle Suche der Zuhörer nach dem Ziel schnell erleichtern, wie unsere Verständnisversuche gezeigt haben. Schließlich wurden die übrigen Sprecher gruppiert und es stellte sich heraus, dass sie öfter überspezifizierten, wenn das überflüssige Adjektiv entropie-reduzierender war als das notwendige. Diese Strategie wurde durch die inzwischen gut etablierte Bevorzugung von Farb- gegenüber Musteradjektiven moduliert und entspricht der begrenzt-rationalen Darstellung.

Zusammenfassend lässt sich sagen, dass unsere Experimente zeigen, dass die Zuhörer auch bei visuell-situierter Kommunikation sensibel auf die Verteilung von Informationen über das Signal hinweg reagieren. Es ist daher naheliegend, dass Rationale Sprecher Redundanzen nutzen, um Informationen gleichmäßiger über ihre Äußerungen zu verteilen um so die Effizienz zu maximieren und gleichzeitig die gemeinsame Anstrengung zu minimieren.

Bibliography

Allopenna, P. D., J. S. Magnuson, and M. K. Tanenhaus
1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4):419 – 439.

Altmann, G. and M. Steedman
1988. Interaction with context during human sentence processing. *Cognition*, 30(3):191 – 238.

Altmann, G. T. and Y. Kamide
1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3):247 – 264.

Anderson, J. R.
1990. *The Rational Analysis of Thought*. Hillsdale, NJ: LEA.

Anderson, J. R.
1991. Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(3):471–485.

Ankener, C. S., M. Sekicki, and M. Staudte
2018. The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology*, 9:2387.

Ariel, M.
1988. Referring and accessibility. *Journal of Linguistics*, 24(1):65–87.

Ariel, M.
2001. Accessibility theory: an overview. In *Text Representation: Linguistic and Physcholinguistic Aspects*, T. Sanders, J. Schilperoord, and W. Spooren, eds., Pp. 29 – 88. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Arnold, J. E.
2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4):495–527.

Arnold, J. E. and S. A. Zerkle
2019. Why do people produce pronouns? pragmatic selection vs. rational models. *Language, Cognition and Neuroscience*, 34(9):1152–1175.

Arts, A., A. Maes, L. Noordman, and C. Jansen
2011a. Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1):361 – 374.

Arts, A., A. Maes, L. Noordman, and C. Jansen
2011b. Overspecification in written instruction. *Linguistics*, 49(3):555—-574.

Arunachalam, S.
2016. A new experimental paradigm to study children's processing of their parent's unscripted language input. *Journal of Memory and Language*, 88:104 – 116.

Aylett, M. and A. Turk
2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Bach, K.
2006. The top 10 misconceptions about implicature. In *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, B. J. Birner and G. L. Ward, eds., Pp. 21–30. Amsterdam: John Benjamins.

Bard, E. and M. Aylett
2004. Referential form, word duration, and modeling the listener in spoken dialogue. In *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, J. Trueswell and M. Tanenhaus, eds., Pp. 173—-191. Cambridge, MA.: MIT Press.

Barr, D., R. Levy, C. Scheepers, and H. Tily
2013. Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Bates, D., M. Maechler, B. Bolker, and S. Walker
2015. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8.

Belke, E.
2006. Visual determinants of preferred adjective order. *Visual Cognition*, 14(3):261–294.

Belke, E. and A. S. Meyer
2002. Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"-"different" decisions. *European Journal of Cognitive Psychology*, 14(2):237–266.

Brennan, S. E.
1995. Centering attention in discourse. *Language and Cognitive Processes*, 10(2):137–167.

Brennan, S. E. and H. H. Clark
1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1482–1493.

Brennan, S. E. and J. E. Hanna
2009. Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2):274–291.

Brodbeck, C., L. Gwilliams, and L. Pylkkänen
2015. Eeg can track the time course of successful reference resolution in small visual worlds. *Frontiers in Psychology*, 6:1787.

Brouwer, H., H. Fitz, and J. Hoeks
2012. Getting real about semantic illusions: Rethinking the functional role of the p600 in language comprehension. *Brain Research*, 1446:127–143.

Brown, P. M. and G. S. Dell
1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19(4):441 – 472.

Brown-Schmidt, S. and J. E. Hanna
2011. Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse*, 2(1):11–33.

Brown-Schmidt, S. and D. Heller
2018. Perspective-taking during conversation. In *Oxford Handbook of Psycholinguistics*, S.-A. Rueschemeyer and M. G. Gaskell, eds. Oxford, UK: Oxford University Press.

Brown-Schmidt, S., S. O. Yoon, and R. A. Ryskin
2015. Chapter three - people as contexts in conversation. volume 62 of *Psychology of Learning and Motivation*, Pp. 59 – 99. Academic Press.

Chase, V. M., R. Hertwig, and G. Gigerenzer
1998. Visions of rationality. *Trends in Cognitive Sciences*, 2(6):206 – 214.

Clark, H. H.
1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H. and S. E. Brennan
1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, L. B. Resnick, L. J. M., and S. D. Teasley, eds., Pp. 127–149. Washington, DC: American Psychological Association.

Clark, H. H. and M. A. Krych
2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1):62 – 81.

Clark, H. H. and C. R. Marshall
1981. Definite knowledge and mutual knowledge. In *Elements of Discourse Understanding*, A. K. Joshi, B. L. Webber, and I. A. Sag, eds., Pp. 10–63. Cambridge, UK: Cambridge University Press.

Clark, H. H. and D. Wilkes-Gibbs
1986. Referring as a collaborative process. *Cognition*, 22(1):1 – 39.

Clarke, A., M. Elsner, and H. Rohde
2013. Where's wally: the influence of visual salience on referring expression generation. *Frontiers in Psychology*, 4:329.

Clarke, A. D. F., M. Elsner, and H. Rohde
2015. Giving good directions: Order of mention reflects visual salience. *Frontiers in Psychology*, 6:1793.

Cooper, R. M.
1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1):84 – 107.

Crain, S. and M. Steedman
1985. *On not being led up the garden path: the use of context by the psychological syntax processor*, Pp. 320–358.

Crocker, M. W., V. Demberg, and E. Teich
2016. Information density and linguistic encoding (ideal). *KI - Künstliche Intelligenz*, 30(1):77–81.

Dambacher, M., R. Kliegl, M. Hofmann, and A. M. Jacobs
2006. Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1):89 – 103.

Davies, C. and J. E. Arnold
2019. Reference and informativeness. In *The Oxford Handbook of Experimental Pragmatics and Semantics*, C. Cummins and N. Katsos, eds., Pp. 29 – 88. Oxford, UK: Oxford University Press.

Davies, C. and 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.

Davies, C., J. Lingwood, and S. Arunachalam
2020. Adjective forms and functions in british english child-directed speech. *Journal of Child Language*, 47(1):159–185.

Degen, J., R. D. Hawkins, C. Graf, E. Kreiss, and N. D. Goodman
2020. When redundancy is useful: A bayesian approach to 'overinformative' referring expressions. *Psychological Review*, Advance online publication.

Dell, G. S.
1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283–321.

Delogu, F., M. Crocker, and H. Drenhaus
2017. Teasing apart coercion and surprisal: Evidence from eye-movements and erps. *Cognition*, 161:46–59.

DeLong, K. A., T. P. Urbach, and M. Kutas
2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117–1121.

Demberg, V. and A. Sayeed
2016. The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLOS ONE*, 11(1):1–29.

Deutsch, W. and T. Pechmann
1982. Social interaction and the development of definite descriptions. *Cognition*, 11(2):159–184.

Dye, M., P. Milin, R. Futrell, and M. Ramscar
2018. Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. *Topics in Cognitive Science*, 10(1):209–224.

Eberhard, K. M., M. J. Spivey-Knowlton, J. C. Sedivy, and M. K. Tanenhaus
1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6):409–436.

Engelhardt, P., S. Demiral, and F. Ferreira
2011. Over-specified referring expressions impair comprehension: An erp study. *Brain and Cognition*, (77):304–314.

Engelhardt, P. E., K. G. Bailey, and F. Ferreira
2006. Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, 54(4):554 – 573.

Engelhardt, P. E., F. Ferreira, and E. G. Patsenko
2010. Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, 63(4):639–645. PMID: 20017058.

Epley, N., B. Keysar, L. V. Boven, and T. Gilovich
2004. Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3):327–339.

Fabiani, M., G. Gratton, D. Karis, and E. Donchin
1987. Definition, identification, and reliability of measurement of the p300 component of the event-related brain potential. *Advances in Psychology*, 2:1 – 78.

Federmeier, K. D. and M. Kutas
1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4):469 – 495.

Fenk-Oczlon, G.
2001. *Familiarity, information flow, and linguistic form*, Pp. 431—-448. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Ferreira, V. S. and G. S. Dell
2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4):296 – 340.

Ferreira, V. S., L. R. Slevc, and E. S. Rogers
2005. How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96(3):263 – 284.

Frank, A. F. and F. T. Jaeger
2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 30*, CogSci08, Pp. 939—-944.

Frank, M. C. and N. D. Goodman
2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Frank, S., L. Otten, G. Galli, and G. Vigliocco
2015. The erp response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1–11.

Frank, S. L.
2010. Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*, CMCL '10, Pp. 81–89, Stroudsburg, PA, USA. Association for Computational Linguistics.

Frank, S. L.
2013. Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, 5(3):475–494.

Fukumura, K.
2018. Ordering adjectives in referential communication. *Journal of Memory and Language*, 101:37 – 50.

Fukumura, K. and R. P. van Gompel
2012. Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model? *Cognitive Science*, 36(7):1289–1311.

Fussell, S. R. and R. M. Krauss
1989. Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6):509–525.

Galati, A. and S. E. Brennan
2010. Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62(1):35 – 51.

Garrett, M. F.
1988. Processes in language production. In *Linguistics: The Cambridge Survey, Biological and Psychological Aspects of Language, vol. 3*, F. Nieumeyer, ed., Pp. 69—-96. Cambridge, UK: Cambridge University Press.

Gatt, A., E. Krahmer, K. van Deemter, and R. P. van Gompel
2017. Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, 41(S6):1457–1492.

Genzel, D. and E. Charniak
2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, Pp. 199–206, Stroudsburg, PA, USA. Association for Computational Linguistics.

Geurts, B. and P. Rubio-Fernández
2015. Pragmatics and processing. *Ratio*, 28(4):446–469.

Gibson, B. S., K. M. Eberhard, and T. A. Bryant
2005. Linguistically mediated visual search: The critical role of speech rate. *Psychonomic Bulletin & Review*, 12(2):276–281.

Gigerenzer, G.
1997. Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics (SJES)*, 133(II):201–218.

Gigerenzer, G. and R. Selten
2001. *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: The MIT Press.

Goodman, N. D. and M. C. Frank
2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.

Grandy, R. E. and R. Warner
2017. Paul grice. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, ed. Metaphysics Research Lab, Stanford University.

Greenhouse, S. and S. Geisser
1959. On methods in the analysis of profile data. *Psychometrika*, 24(2):95–112.

Grice, P.
1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Grice, P. H.
1975. Logic and conversation. In *Syntax and Semantics Vol 3, Speech Acts*, P. Cole and J. L. Morgan, eds., Pp. 41 – 58. New York: Academic Press.

Grodner, D. and J. C. Sedivy
2011. *The Effect of Speaker-Specific Information on Pragmatic Inferences*, Pp. 239–272.

Gundel, J. K., N. Hedberg, and R. Zacharski
1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

Hale, J.
2001. A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, NAACL '01, Pp. 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hale, J.
2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 2(32):101–123.

Hale, J.
2006. Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4):643–672.

Hale, J.
2016. Information-theoretical complexity metrics. *Language and Linguistics Compass*, 10(9):397–412.

Hauk, O.
2005. Prediction in language comprehension and production [special issue]. *Language, Cognition and Neuroscience*, 31(1).

Heller, D., K. S. Gorman, and M. K. Tanenhaus
2012. To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, 4(2):290–305.

Hoeks, J. C. J., L. A. Stowe, P. Hendriks, and H. Brouwer
2013. Questions left unanswered: How the brain responds to missing information. *PLOS ONE*, 8(10):1–9.

Horn, L.
1984. *Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature*, Pp. 11–42. Georgetown University Press.

Horn, L.
2004. *Implicature*, Pp. 3–28. Oxford, UK: Blackwell Publishing.

Horton, W. S. and R. J. Gerrig
2005. The impact of memory demands on audience design during language production. *Cognition*, 96(2):127 – 142.

Horton, W. S. and B. Keysar
1996. When do speakers take into account common ground? *Cognition*, 59(1):91 – 117.

Huettig, F., J. Rommers, and A. S. Meyer
2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2):151 – 171. Visual search and visual world: Interactions among visual attention, language, and working memory.

Isaacs, E. A. and H. H. Clark
1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1):26–37.

Jaeger, T.
2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Just, M. A. and P. Carpenter
1993. The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale*, 47 2:310–39.

Kamide, Y., G. T. Altmann, and S. L. Haywood
2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1):133 – 156.

Keysar, B., D. J. Barr, and W. S. Horton
1998. The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7(2):46–49.

Keysar, B. and W. S. Horton
1998. Speaking with common ground: from principles to processes in pragmatics: a reply to polichak and gerrig. *Cognition*, 66(2):191 – 198.

Knoeferle, P.
2015. *Visually Situated Language Comprehension in Children and in Adults*, Pp. 57–75. New Delhi: Springer India.

Knoeferle, P., M. W. Crocker, C. Scheepers, and M. J. Pickering
2005. The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, 95(1):95 – 127.

Knoeferle, P. and E. Guerra
2016. Visually situated language comprehension. *Language and Linguistics Compass*, 10(2):66–82.

Knoeferle, P. and H. Kreysa
2012. Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, 3:538.

Koolen, R., A. Gatt, M. Goudbeek, and E. Krahmer
2011. Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13):3231 – 3250.

Koolen, R., M. Goudbeek, and E. Krahmer
2013. The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37(2):395–411.

Koolen, R., E. Krahmer, and M. Swerts
2016. How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science*, 40(7):1617–1647.

Krauss, R. M. and S. Glucksberg
1977. Social and nonsocial speech. *Scientific American*, 236(2):100–105.

Krauss, R. M. and S. Weinheimer
1964. Changes in reference phrases as a function of frequency of usage in social interaction: a preliminary study. *Psychonomic Science*, 1(1):113–114.

Kuperberg, G. R., A. Choi, N. Cohn, M. Paczynski, and R. Jackendoff
2010. Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, 22(12):2685–2701.

Kutas, M. and K. D. Federmeier
2009. N400. *Scholarpedia*, 4(10):7790. revision #91546.

Kutas, M. and 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(1):621–647.

Lawrence, M. A.
2011. *ez: Easy analysis and visualization of factorial experiments*. R package version 3.0-0.

Levelt, W. J. M.
1989. *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J. M. and A. S. Meyer
2000. Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, 12(4):433–452.

Levelt, W. J. M., A. Roelofs, and A. S. Meyer
1999. A theory of lexical access in speech production. *Behavioural and Brain Sciences*, (22):1–75.

Levinson, S. C.
1983. *Pragmatics*. Cambridge University Press.

Levinson, S. C.
2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press.

Levy, R.
2008. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Levy, R. P. and T. F. Jaeger
2007. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems 19*, B. Schoelkopf, J. C. Platt, and T. Hoffman, eds., Pp. 849–856. MIT Press.

Linzen, T. and T. F. Jaeger
2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, 40(6):1382–1411.

Maes, A., A. Arts, and L. Noordman
2004. Reference management in instructive discourse. *Discourse Processes*, 37(2):117–144.

Magliero, A., T. R. Bashore, M. G. Coles, and E. Donchin
1984. On the dependence of p300 latency on stimulus evaluation processes. *Psychophysiology*, 21(2):171–186.

Mahowald, K., E. Fedorenko, S. T. Piantadosi, and E. Gibson
2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313 – 318.

Marr, D.
1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt and Co., Inc.

Marshall, S. P.
2000. Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity.

Marshall, S. P.
2002. The index of cognitive activity: measuring cognitive workload. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, Pp. 7–7.

Matin, E., K. C. Shao, and K. R. Boff
1993. Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4):372–380.

Meyer, A. S.
1996. Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, 35(4):477 – 496.

Meyer, A. S., A. Roelofs, and L. Brehm
2019. Thirty years of speaking: An introduction to the special issue. *Language, Cognition and Neuroscience*, 34(9):1073–1084.

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

Nieuwland, M. and J. Van Berkum
2008a. The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from erps. *Brain and Language*, 106(2):119–131.

Nieuwland, M. and J. Van Berkum
2008b. The neurocognition of referential ambiguity in language comprehension. *Language and Linguistic Compass*, 2(4):603–630.

Noveck, I. A. and A. Reboul
2008. Experimental pragmatics: a gricean turn in the study of language. *Trends in Cognitive Sciences*, 12(411):425–431.

Olson, D. R.
1970. Language and thought: aspects of a cognitive theory of semantics. *Psychological review*, 77(4):257–73.

Paraboni, I., K. van Deemter, and J. Masthoff
2007. Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33(2):229–254.

Pechmann, T.
1989. Incremental speech production and referential overspecification. *Linguistics*, 27(1):89–110.

Piantadosi, S. T., H. Tily, and E. Gibson
2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Pickering, M. J. and S. Garrod
2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, (27):169–190.

Pogue, A., C. Kurumada, and M. K. Tanenhaus
2016. Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*, 6:2035.

Qing, C., D. Lassiter, and J. Degen
2018. What do eye movements in the visual world reflect? a case study from adjectives. In *CogSci*.

R Core Team
2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reali, F., M. J. Spivey, M. J. Tyler, and J. Terranova
2006. Inefficient conjunction search made efficient by concurrent spoken delivery of target identity. *Perception & Psychophysics*, 68(6):959–974.

Richardson, D. C., R. Dale, and J. M. Tomlinson
2009. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33(8):1468–1482.

Roßnagel, C.
2000. Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30(3):429–445.

Rubio-Fernández, P.
2008. On the automaticity of egocentricity: A review of the egocentric anchoring and adjustment model of perspective taking. *UCL Working Papers in Linguistics*, 20:247 – 274.

Rubio-Fernández, P.
2020. Redundant color words are more efficient than shorter descriptions.

Rubio-Fernández, P.
2016. How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7:1–15.

Rubio-Fernández, P.
2019. Overinformative speakers are cooperative: Revisiting the gricean maxim of quantity. *Cognitive Science*, 43(11):e12797.

Ryskin, R., C. Kurumada, and S. Brown-Schmidt
2019. Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, 43(8):e12769.

Saryazdi, R., J. Bannon, and C. G. Chambers
2019. Age-related differences in referential production: A multiple-measures study. *Psychology and Aging*, 34(6):791–804.

Sassenhagen, J., M. Schlesewsky, and I. Bornkessel-Schlesewsky
2014. The p600-as-p3 hypothesis revisited: Single-trial analyses reveal that the late eeg positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137:29 – 39.

Scheepers, C. and M. W. Crocker
2004. Constituent order priming from reading to listening: a visual-world study. In *The on-line study of sentence comprehension: Eyetracking, ERP and beyond*, P. 167–185, New York.

Schmitt, B. M., A. S. Meyer, and W. J. M. Levelt
1999. Lexical access in the production of pronouns. *Cognition*, 69(3):313–335.

Schober, M. F. and H. H. Clark
1989. Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2):211 – 232.

Schriefers, H.
1993. Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19:841 – 850.

Schriefers, H. and T. Pechmann
1988. Incremental production of noun-phrases by human speakers. In *Advances in natural language generation*, M. Zock and G. Sabah, eds., P. 172–179. London: Pinter.

Sedivy, J., M. Tanenhaus, C. Chambers, and G. Carlson
1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.

Sedivy, J. C.
2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32(1):3–23.

Sedivy, J. C.
2005. Evaluating explanations for referential context effects: Evidence for gricean mechanisms in online language interpretation. In *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, J. C. Trueswell and M. K. Tanenhaus, eds., Pp. 345—-364. Cambridge, MA: MIT Press.

Sekicki, M. and M. Staudte
2018. Eye'll help you out! how the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, 42(8):2418–2458.

Shannon, C. E.
1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Sikos, L., S. Tomlinson, C. Heins, and D. Grodner
2019a. What do you know? erp evidence for immediate use of common ground during online reference resolution. *Cognition*, 182:275–285.

Sikos, L., N. Venhuizen, H. Drenhaus, and M. Crocker
2019b. Reevaluating pragmatic reasoning in web-based language games.

Simon, H. A.
1955. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Smith, N. and R. Levy
2013. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319.

Sperber, D. and D. Wilson
1986. *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Spivey, M. J., M. J. Tyler, K. M. Eberhard, and M. K. Tanenhaus
2001. Linguistically mediated visual search. *Psychological Science*, 12(4):282–286.

Tanenhaus, M., M. Spivey-Knowlton, K. Eberhard, and J. Sedivy
1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634.

Tanenhaus, M. K., J. S. Magnuson, D. Dahan, and C. Chambers
2000. Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29(6):557–580.

Tarenskeen, S., M. Broersma, and B. Geurts
2015. Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, 6:1703.

Tourtouri, E., F. Delogu, and M. W. Crocker
2015. Erp indices of situated reference in visual contexts. In *CogSci*.

Tourtouri, E. N., F. Delogu, L. Sikos, and M. W. Crocker
2019. Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, 3(2):175–202.

Treisman, A. M. and G. Gelade
1980. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97 – 136.

Van Berkum, J. J. A., C. M. Brown, P. Zwitserlood, V. Kooijman, and P. Hagoort
2005. Anticipating upcoming words in discourse: Evidence from erps and reading times. *Journal of Experimental Psychology Learning Memory and Cognition*, 31(3):443–467.

van Gompel, R. P. G., K. van Deemter, A. Gatt, R. Snoeren, and E. J. Krahmer
2019. Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, 126(3):345–373.

Vanlangendonck, F., R. M. Willems, L. Menenti, and P. Hagoort
2016. An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, 31(6):741–750.

Venhuizen, N. J., M. W. Crocker, and H. Brouwer
2019. Semantic entropy in language comprehension. *Entropy*, 21(12):1159.

Vogels, J.
2014. *Referential choices in language production: The role of accessibility*. PhD thesis, Tilburg center for Cognition and Communication (TiCC), Tilburg.

Vogels, J., V. Demberg, and J. Kray
2018. The index of cognitive activity as a measure of cognitive processing load in dual task settings. *Frontiers in Psychology*, 9:2276.

Vogels, J., D. M. Howcroft, E. N. Tourtouri, and V. Demberg
2019. How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience*, 0(0):1–15.

Vogels, J., E. Krahmer, and A. Maes
2015. How cognitive load influences speakers' choice of referring expressions. *Cognitive Science*, 39(6):1396–1418.

Vonk, W., L. G. Hustinx, and W. H. Simons
1992. The use of referential expressions in structuring discourse. *Language and Cognitive Processes*, 7(3-4):301–333.

Wardlow Lane, L. and V. S. Ferreira
2008. Speaker-external versus speaker-internal forces on utterance form: do cognitive demands override threats to referential success? *Journal of experimental psychology. Learning, memory, and cognition*, 34(6):1466—1481.

Weber, A., M. Grice, and M. W. Crocker
2006. The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, 99(2):B63 – B72.

Wolfe, J. M.
1994. Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238.

Wolfe, J. M.
1998. What can 1 million trials tell us about visual search? *Psychological Science*, 9(1):33–39.

Wu, S., A. Bachrach, C. Cardenas, and W. Schuler
2010. Complexity metrics in an incremental right-corner parser. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Pp. 1189–1198, Uppsala, Sweden. Association for Computational Linguistics.

Yoon, S. O., S. Koh, and S. Brown-Schmidt
2012. Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, 19(4):699–707.

Zipf, G. K.
1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison - Wesley Press.

Appendix A

Object images

The object images presented below were used to create the visual stimuli in all experiments. In Experiment 1, objects appeared in all four colours (red, blue, green, yellow) and three patterns (checkered, dotted, striped). Experiments 2-5 used only three of the colours (red, blue, green) and all of the patterns.

For the visual stimuli in Experiment 1, we used the images of all 32 objects.

For the visual stimuli in Experiments 2 and 3, the images of 30 objects were used; 'watch' (*Armbanduhr*) and 'blazer' (*Blazer*) were not included.

For the visual stimuli in Experiments 4 and 5 we used the images for: 'ball' (*Ball*), 'bucket' (*Eimer*), 'flip-flop' (*Flip-flop*), 'watering can' (*Gießkanne*), 'belt' (*Gürtel*), 'suitcase' (*Koffer*), 'tie' (*Krawatte*), 'lamp' (*Lampe*), 'paper cup' (*Pappbecher*), 'rucksack' (*Rucksack*), 'umbrella' (*Schirm*), 'bowl' (*Schüssel*), 'shoe' (*Schuh*), 'sunglasses' (*Sonnenbrille*), 'boot' (*Stiefel*), 'mug' (*Tasse*), 'mitt' (*Topfhandschuh*), 'clock' (*Wanduhr*).

In what follows, all 32 object images are presented in alphabetical order in German.

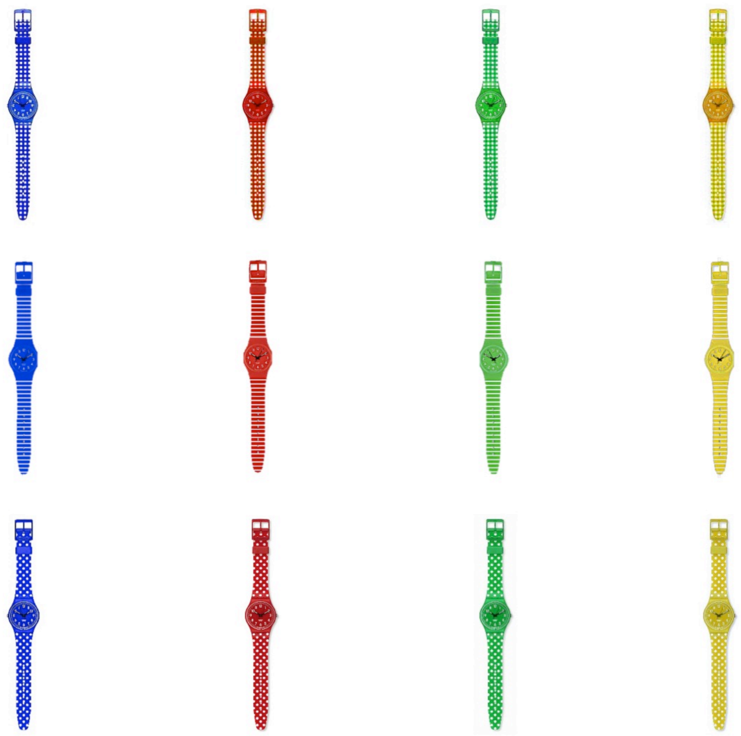


Figure A.1 Armbanduhr

Figure A.2 Ball

Figure A.3 Blazer

Figure A.4 Eimer

Figure A.5 Flip-flop

Figure A.6 Geschenkband

Figure A.7 Gießkanne

Figure A.8 Gürtel

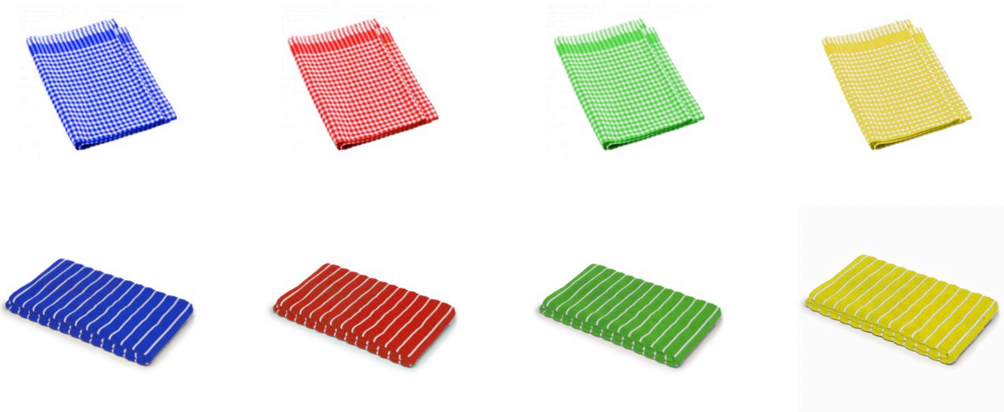


Figure A.9 Handtuch

Figure A.10 Hemd

Figure A.11 Kerze

Figure A.12 Kissen

Figure A.13 Kleid

Figure A.14 Koffer

Figure A.15 Krawatte

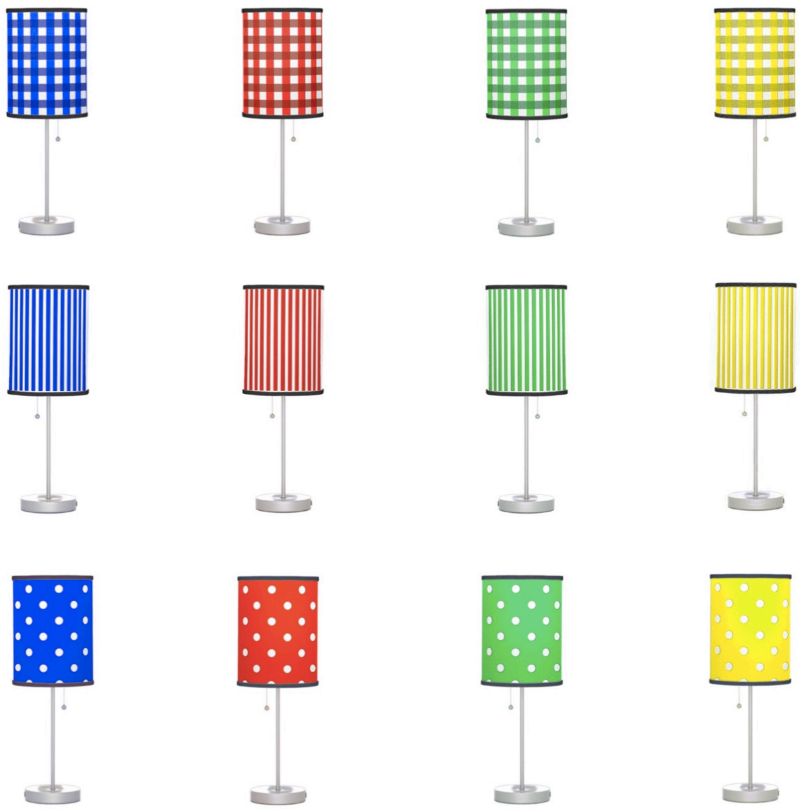


Figure A.16 Lampe

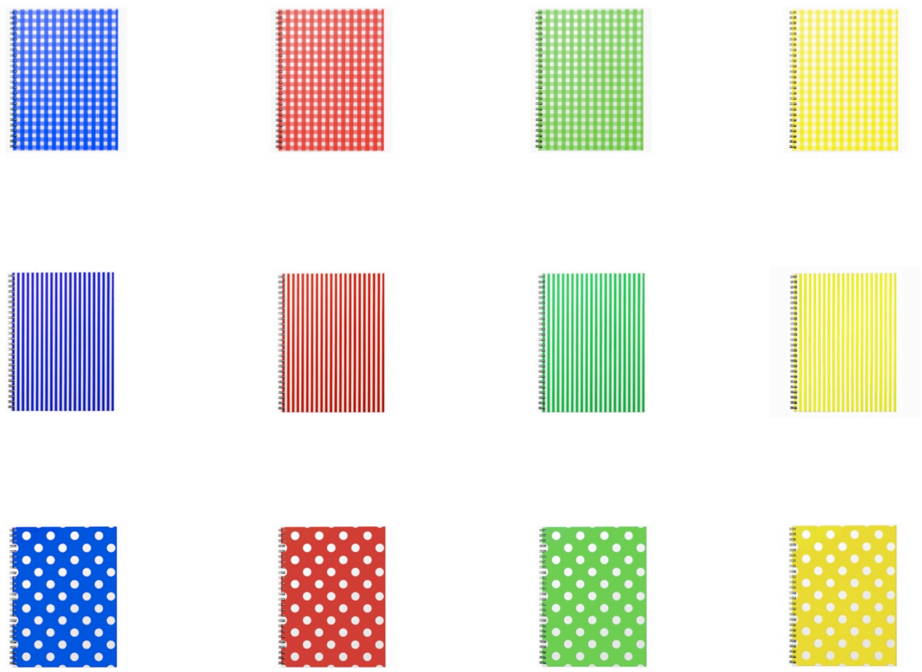


Figure A.17 Notizblock

Figure A.18 Pappbecher

Figure A.19 Rucksack

Figure A.20 Schirm

Figure A.21 Schuh

Figure A.22 Schürze

Figure A.23 Schüssel

Figure A.24 Sessel

Figure A.25 Sofa

Figure A.26 Sonnenbrille

Figure A.27 Stiefel

Figure A.28 Tasche

Figure A.29 Tasse

Figure A.30 Topfhandscuh

Figure A.31 T-shirt

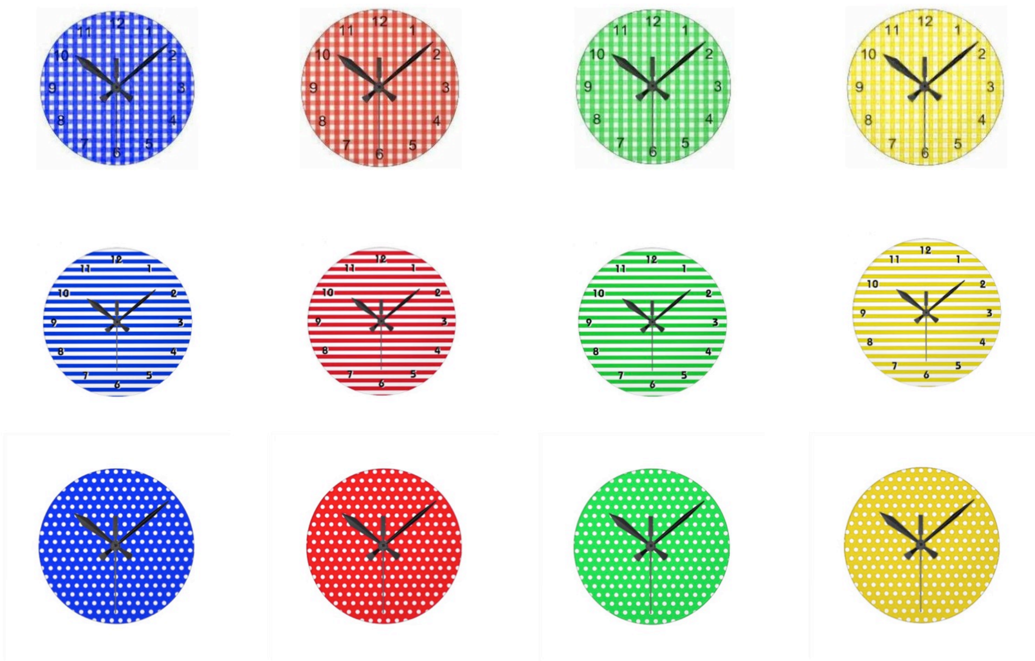


Figure A.32 Wanduhr

Appendix B

Pattern visual stimuli

Figure B.1 Experiment 1. Sample visual scenes for a pattern experimental item, paired with the spoken instruction 'Find the dotted bowl'.

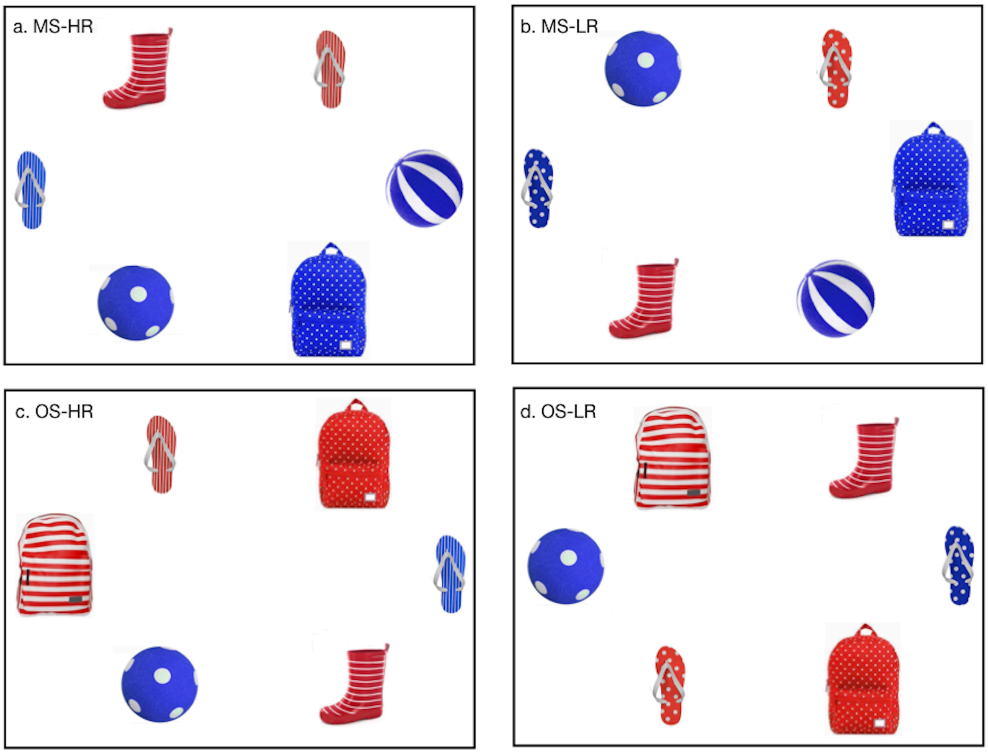


Figure B.2 Experiments 2 and 3. Sample visual scenes for a pattern experimental item, paired with the spoken instruction 'Find the dotted ball'.

Appendix C

Task prompt

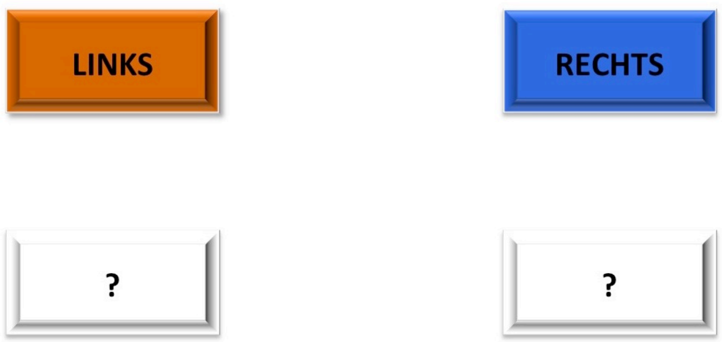


Figure C.1 Experiments 1-3. Screen prompting participants to perfrom the task

ELLI N. TOURTOURI

Curriculum Vitae

Contact Address
Max Planck Institute for Psycholinguistics
Wundtlaan 1
6525 XD Nijmegen
The Netherlands

E-mail
elli.tourtouri@mpi.nl

Phone
+31 24 3521379

EDUCATION

PhD in Psycholinguistics, Saarland University — Apr 2020
Saarbrücken, Germany
Advisor: Prof. Matthew W. Crocker
Thesis title: *Rational Redundancy in Situated Communication*

MA in Linguistics (Research), Utrecht University — Aug 2012
Specialisation: Psycholinguistics
Utrecht, The Netherlands
Thesis title: Does personality modulate brain responses to affective speech?

BA (Hons) in Greek Philology, Aristotle University of Thessaloniki — July 2010
Specilisation: Linguistics
Thessaloniki, Greece

ACADEMIC POSITIONS

Postdoctoral researcher — Apr 2020 –
Psychology of Language Department
Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

Doctoral researcher — Oct 2014 – Dec 2018
SFB 1102 "Information Density and Linguistic Encoding" (iDeaL)
(Funded by the German Research Council – DFG)
Saarbrücken, Germany

Lecturer — Apr 2013 – Sept 2014
Department of Computational Linguistics and Phonetics
Saarland University
Saarbrücken, Germany

Doctoral researcher — Jan 2013 – Sept 2014
Department of Computational Linguistics and Phonetics
Saarland University
Saarbrücken, Germany

Intern — Mar – June 2012
"Smiling Project: How audible smiles affect speech comprehension"
Utrecht Institute of Linguistics OTS
Utrecht, The Netherlands

PUBLICATIONS

In prep. **Tourtouri E. N.**, Delogu F., & Crocker M. W. *Rational ERPs: The Comprehension of Over-specification in Visually Situated Comprehension.*

2020. **Tourtouri E. N.** *Rational Redundancy in Situated Communication.* PhD thesis.

2019. **Tourtouri E. N.**, Delogu F., Sikos L., & Crocker M. W. Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science.* [Special issue on Speakers and Listeners in the Visual World].

2019. Vogels J., Howcroft D. M., **Tourtouri E. N.**, & Demberg V. How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience.*

2017. **Tourtouri E. N.**, Delogu F., & Crocker M. W. Specificity and Entropy Reduction in situated referential processing. In Gunzelmann G., Howes A., Tenbrink T., & Davelaar E. J. (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, (pp. 3356-3361). Austin, TX: Cognitive Science Society.

2015. **Tourtouri E. N.**, Delogu F., & Crocker M. W. ERP Indices of Situated Reference in Visual Contexts. In Noelle D. C., Dale R., Warlaumont A. S., Yoshimi J., Matlock T., Jennings C. D., & Maglio P. P. (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, (pp. 2422-2427). Austin, TX: Cognitive Science Society.

PRESENTATIONS

2020. Gregory Scontras, Zeinab Kachakeche, Autin Nguyen, Cesar Rosales, Suttera Samonte, Einat Shetreet, Yuxin Shi, **Elli Tourtouri** & Nitzan Trainin. Cross-linguistic evidence for subjectivity-based adjective ordering preferences. Accepted at TExMod2020, Tuebingen, Germany (Conference postponed).

2019. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Differential indices of comprehension in visual search: Evidence from over-specification. Talk presented at the Embodied and Situated Language Processing (ESLP) / Attentive Listener in the Visual World (AttLis) joint conference. August 28-30, Berlin, Germany.

2019. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Rational ERPs: The comprehension of over-specification in visually-situated contexts. Talk presented at the 8th Biannual Experimental Pragmatics (XPrag) Conference. June 19-21, Edinburgh, UK.

2019. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Understanding over-specification: A visual-world ERP study. Poster presented at CUNY 2019. March 29-31, Boulder, CO, USA.

2018. **Tourtouri, E. N.**, Sikos, L., & Crocker, M. W. Referential Overspecification: From Egocentricity to Rationality. Poster presented at AMLaP 2018. September 6-9, Berlin, Germany.

2018. **Tourtouri, E. N.**, Sikos, L., & Crocker, M. W. Referential overspecification as a rational strategy. Poster at the 10th International Workshop on Language Production (IWLP). July 2-4, Nijmegen, The Netherlands.

2018. **Tourtouri, E. N.**, Sikos, L., & Crocker, M. W. Referential entropy influences the production of overspecifications. Poster to be presented at the X. Dubrovnik Conference on Cognitive Science Communication, Pragmatics, and Theory of Mind (DuCog). May 24-27, Dubrovnik, Croatia.

2018. **Tourtouri, E. N.**, Sikos, L., & Crocker, M. W. Referential Entropy influences Overspecification: Evidence from Production. Poster presented at the 31st Annual CUNY Sentence Processing Conference. March 15-17, Davis CA, USA.

2018. Vogels, J., Howcroft, D., **Tourtouri, E. N.**, & Demberg, V. Do speakers adapt object descriptions to listeners under load? Talk presented at the 31st Annual CUNY Sentence Processing Conference. March 15-17, Davis CA, USA.

2017. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. The interplay of specificity and referential entropy reduction in situated communication. Talk presented at the 10th Annual Embodied and Situated Language (ESLP) Conference. September 10-12, Moscow, Russia.

2017. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Specificity and Entropy Reduction in situated referential processing. Poster presented at the 39th Annual Conference of the Cognitive Science Society. July 26-29, London, UK.

2017. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Overspecification and uniform reduction of visual entropy facilitate referential processing. Poster presented at the 7th Biannual Experimental Pragmatics (XPrag) Conference. June 21-23, Cologne, Germany.

2017. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. Overspecifications efficiently manage referential entropy in situated communication. Talk presented at the 39th Annual Conference of the German Linguistic Society (DGfS). March 8-10, Saarbrücken, Germany.

2017. **Tourtouri, E. N.**, Specificity and entropy reduction in situated referential processing. Invited talk at the Experimental Linguistics Talks Utrecht (ELiTU). February 6, Utrecht, Netherlands.

2015. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. ERP indices of situated reference in visual contexts. Poster presented at the 37th Annual Conference of the Cognitive Science Society. July 23-25, Los Angeles, USA.

2015. **Tourtouri, E. N.**, Delogu, F., & Crocker, M. W. ERP indices of referential informativity in visual contexts. Poster presented at the 28th Annual CUNY Sentence Processing Conference. March 19-21, Los Angeles, USA.