How visual saliency affects referent accessibility

Jorrig Vogels (j.vogels@uvt.nl)
Emiel Krahmer (e.j.krahmer@uvt.nl)
Alfons Maes (maes@uvt.nl)
Tilburg centre for Cognition and Communication (TiCC), Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

As referents are more accessible in discourse, they can be referred to with more attenuated expressions, such as pronouns. Accessibility is known to be affected by the referent's saliency in the linguistic context, but much less is known about the effect of saliency in the visual context. In this paper, we investigate whether a referent's visual saliency affects the choice of referring expression in a discourse context. The results of a story completion experiment show that visually salient referents induce more attenuated expressions, but only when they are linguistically non-salient. Linguistically salient referents receive more reduced references when they are visually *non*-salient. We argue that visual saliency affects accessibility when the impact of linguistic factors is moderated. In addition, when the story does not match people's expectations, processing difficulties might result in the use of less costly expressions.

Keywords: saliency; accessibility; visual context; referring expressions; language production.

Introduction

In discourse, the same entity can be referred to with different types of expressions, ranging from elaborate descriptions involving full noun phrases and modifiers (e.g. *the blonde girl with the big earrings*) to short, low-informative elements such as pronouns (e.g. *she*). It has been argued that the referring expression a speaker chooses reflects the cognitive status of the referent (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993). That is, it is believed that speakers make assumptions about the degree of activation of the referent in the memory of their addressees. The more accessible an entity is to the addressee, the less information a referring expression needs to contain to be correctly understood. In addition, production processes are constrained by speaker-internal factors, such as cognitive load, that may affect referent accessibility in the speaker's memory (e.g. Arnold & Griffin, 2007).

An important source of a referent's accessibility is its saliency. The role of saliency in the choice of referring expressions within a discourse has mainly been investigated in relation to the preceding linguistic context. For example, subjects and topics (i.e. what a sentence is about) are considered to be salient entities in a discourse (e.g. Arnold, 1998; Givón, 1983; Gordon, Grosz, & Gilliom, 1993; Grosz, Joshi, & Weinstein, 1995). Hence, a referent that is the subject or the topic of a preceding utterance is more accessible than other possible referents. Therefore, it is more likely to be referred to with an attenuated expression, such as a pronoun, while a less accessible referent needs a more elaborate description, such as a full noun phrase. Other factors that affect a referent's saliency in the linguistic context include recency, givenness, thematic roles and syntactic position (e.g. Clark & Sengul, 1979; Gundel et al., 1993; Stevenson, Crawley, & Kleinman, 1994).

It is much less clear whether saliency in the *visual* context also plays a role in the accessibility of referents in a discourse. In everyday life, the things we talk about are often not only available to us through previous linguistic mention, but in many cases they are also physically present. In this paper, we investigate whether the accessibility of a referent is influenced by its salience in the visual context. Language production studies that have taken into account the visual context suggest that speakers use non-linguistic information in planning their utterances. For instance, the syntactic structure of visual scene descriptions is affected by where people look in the scene (e.g. Gleitman et al., 2007). Visual information is also used to choose a referring expression. For example, people use disambiguating expressions to refer to visually ambiguous referents (e.g. Brown-Schmidt & Tanenhaus, 2006; Ferreira, Slevc, & Rogers, 2005), and more reduced expressions when referents are visually in focus (Beun & Cremers, 1998).

When a referent is also accessible from the linguistic context, it might be the case that effects of visual information are overruled by linguistic information. In a story completion experiment, Arnold and Griffin (2007) found that participants used fewer pronouns to refer to the target character when a competitor referent was mentioned in the linguistic context. Whether the competitor was also visually present in the target scene did not make a difference, suggesting that the linguistic presence of the competitor affected accessibility, but not its visual presence. In contrast, Fukumura, Van Gompel and Pickering (2010) found in a similar experiment that visual context did influence the choice of referring expression: Participants used fewer pronouns to refer to the target referent when a competitor was visually present than when it was not visually present. However, the effect of the visual context was larger when the competitor was also linguistically present than when it was not mentioned at all. This suggests that accessibility is affected most by linguistic context, but that the influence of visual context becomes more apparent when the linguistic context is less compelling.

Fukumura et al. (2010) argued that the visual presence of the competitor in their experiment reduced the salience of the referent, which led to a decrease in accessibility.

However, it is not clear whether this is really an effect of salience, or merely an interference effect due to the fact that multiple possible referents have to be kept in memory. Therefore, we argue that, instead of varying the number of competing entities, the perceptual prominence of the referent itself should be taken into account. A number of properties have been identified as important cues to perceptual salience, such as size, centrality, color, foregrounding, orientation, intensity and visual complexity (e.g. Coco & Keller, 2009; Kelleher, Costello, & Van Genabith, 2005; Mazza, Turatto, & Umiltà, 2005; Parkhurst, Law, & Niebur, 2002). Since the role of these properties in determining the accessibility of a referent in discourse is still unclear, it remains an open question in what way visual salience affects the choice of a referring expression in interaction with linguistic context.

In this paper, we present a story completion experiment in which we investigate the effect of a referent's visual salience on the use of pronouns versus full noun phrases in Dutch narrative discourse. Since Dutch has a distinction between full and reduced pronouns, we also examine the use of reduced pronouns versus full pronouns (cf. Kaiser & Trueswell, 2004). If visually salient characters are more accessible, they are expected to be referred to with more reduced expressions than visually non-salient entities. Thus, we predict more pronouns than full NPs and more reduced pronouns than full pronouns in references to visually salient referents. In addition, if linguistic information is more important in determining accessibility than visual information, as suggested by previous studies, an effect of visual saliency should at least be expected in contexts where linguistic saliency is moderated.

Methods

Participants

Sixty-four students (54 female; mean age 21) from Tilburg University participated for course credit. They were all native speakers of Dutch and had normal or corrected to normal vision.

Materials

Sixteen short stories served as the stimulus items. Each item consisted of two pictures, two context sentences and the onset of a third sentence, which had to be completed by the participants. The first two context sentences accompanied the first picture of a pair, while the onset of the third sentence was aligned with the onset of the second picture. The pictures showed a male and a female character in a certain situation. One character was the target referent, which always performed an action in the second picture. Therefore, it was expected to be mentioned in subject position in the participant's completion. We manipulated the visual and the linguistic salience of the target referent, resulting in four different picture pairs for each stimulus item. These are exemplified in Figure 1.

The target referent was either mentioned as the subject of the second context sentence, which directly preceded the sentence that had to be completed by the participants (condition A & B in Figure 1), or as the subject of the first context sentence (condition C & D in Figure 1). In the former case, the target referent was considered linguistically salient. In the latter case, it was considered linguistically non-salient. This is in line with the common assumption in theories of reference that the subject or topic of the preceding utterance is the most prominent entity at the start of the current utterance (e.g. Grosz et al., 1995). When the target referent was the subject of the first context sentence, the other character was the subject of the second sentence, and vice versa. This subject shift was included to ensure that neither character became so linguistically salient that any effects of visual salience would be overruled.

For each item, the linguistic context was the same in all versions of the picture pairs. The first context sentence always started with the phrase *Er was eens* 'Once upon a time there was', followed by an indefinite subject, which referred to the female character (either *een vrouw* 'a woman' or *een meisje* 'a girl') in half of the cases and to the male character in the other half (either *een man* 'a man' or *een jongen* 'a boy'). The subject was modified by a relative clause describing the situation (e.g. *die een gesprek voerde* 'who had a conversation'), always followed by a prepositional phrase introducing the other character (e.g. *met een jongen* 'with a boy'). Subsequently, this character became the subject of the second sentence, which described a physical or emotional state (e.g. *De jongen raakte enorm verveeld* 'The boy got really bored'). The adjective used here always denoted a temporary, event-like property, such as *verveeld* 'bored', which would make it less likely that the second picture would be described as a habitual or generic event. To further emphasize the episodic nature of the stories, the finite verb in the second sentence was always a dynamic verb, such as *worden* 'to become'. The onset of the third sentence always consisted of the word *Daarom* 'That's why'. Because Dutch is a verb second language, this means that participants had to start their utterance with a finite verb, directly followed by the subject, which was the constituent of interest. All sentences were recorded by a female native speaker of Dutch. A pretest of the sentences revealed that three items contained a bias for continuing the context sentences with either one or the other character. After the sentences were adapted, the bias disappeared.

In the pictures, the target referent either appeared in a central position in the foreground (condition A & C in Figure 1), or in a more peripheral position in the background (condition B & D in Figure 1). In the former case, the target referent was considered visually salient, while in the latter case it was considered visually non-salient. Since the other character was in the background when the target referent was in the foreground and vice versa, visual salience was always relative to the other character. In most cases, the foregrounded character also partly occluded the backgrounded character. Some additional steps were taken to emphasize the difference in visual salience. Firstly, the character in the foreground was made more prominent by putting a spotlight and the camera's focus on this person. Secondly, the positions of the two characters were kept constant across items, such that the distance between them was always the same. In addition, the action in the second picture always involved at least standing up from a chair, causing the target referent to be upright at all times. To minimize distraction from the two characters caused by other objects, the only furniture used were two chairs and an optional table, and photographing was done against a white screen. Four couples posed for all pictures. To avoid any effects of the left-to-right orientation of the characters in the pictures, a mirror version was created for each picture pair (not shown in Figure 1).

In the first picture of each story, both characters were in a neutral position (e.g. sitting next to each other). In the second picture, either the male or the female character performed a simple action, which was one of two kinds: Either getting an object related to the state of the character described in the second sentence (e.g. getting a pillow when tired), or walking away. Care was taken that the action depicted in the second picture was compatible with the context sentences in the different versions of an item, i.e. both when the man and when the woman was the agent. For example, the action of getting a beer in reaction to the man being thirsty can be performed by both characters, since one can do this for oneself or for someone else.

An additional 20 items serving as fillers and 4 practice items were constructed. These were similar to the experimental items, except that 5 items included only one character and another 9 items included two characters of the same gender. In addition, the characters sometimes had roles like 'a teacher' or 'a saleswoman'. The filler and

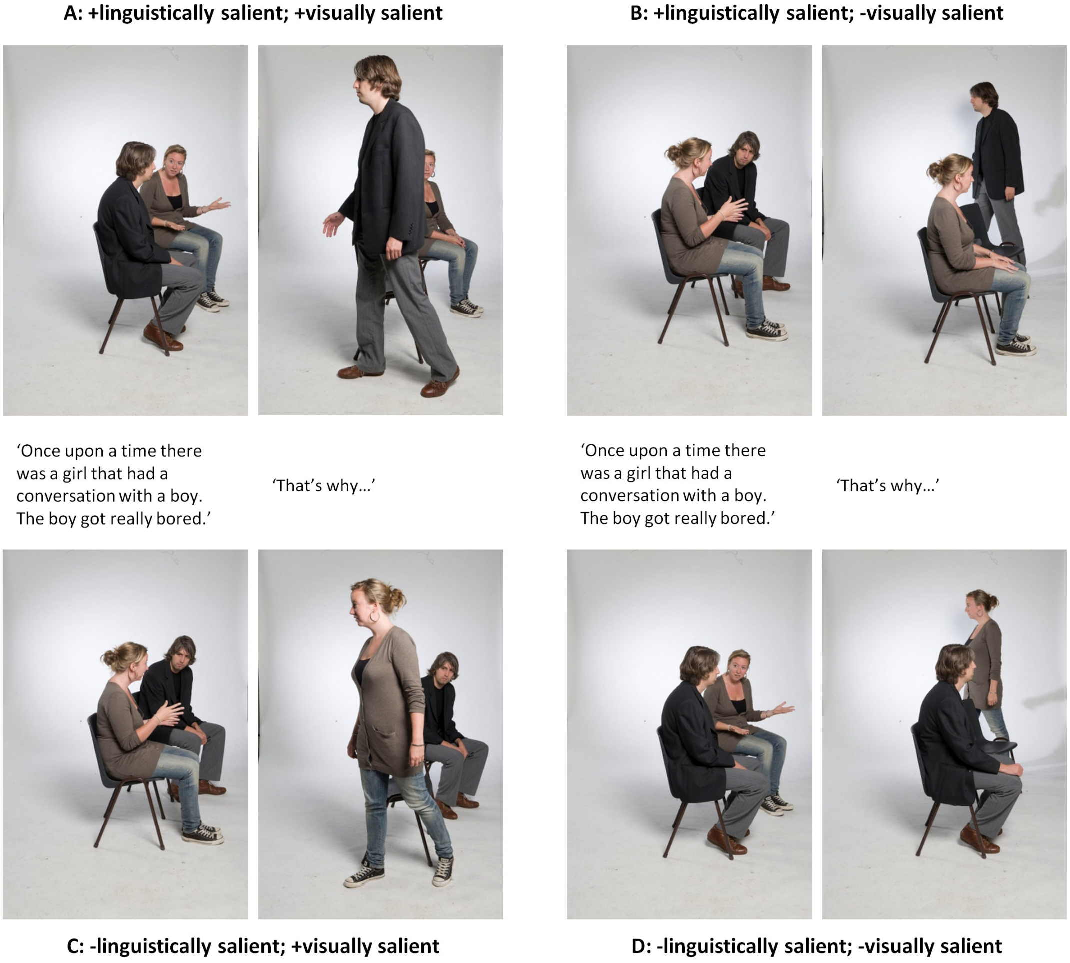
A: +linguistically salient; +visually salient
B: +linguistically salient; -visually salient
'Once upon a time there was a girl that had a conversation with a boy. The boy got really bored.'
'That's why…'
'Once upon a time there was a girl that had a conversation with a boy. The boy got really bored.'
'That's why…'
C: -linguistically salient; +visually salient
D: -linguistically salient; -visually salient

Figure 1: A stimulus item in four different conditions: (A) target referent (i.e. the person performing the action in the second picture) is both linguistically and visually salient; (B) target referent is linguistically but not visually salient; (C) target referent is visually but not linguistically salient; (D) target referent is neither linguistically nor visually salient. The corresponding context sentences are translations of the Dutch originals.

practice items came in only one version. All items were distributed over eight lists using a Latin square design, such that each list contained one version of a given stimulus item. On each list, items were quasi-randomized, with the filler items having a fixed position and no two experimental items occurring in consecutive slots.

Procedure

Participants sat in a low noise cabin behind a computer screen. In front of the computer screen was a microphone to record the participants' responses. The experiment was assembled and run with the E-Prime 2.0 software program. Participants were instructed to complete each story initiated by the context sentences in such a way that it would fit in with the situation shown in the second picture. They were told that they had to build a sentence that connected to the word *Daarom* 'That's why'. They were not allowed to repeat this word, because this would cause a break in the continuation of the story. Participants were further instructed to use their first intuitions about how to complete the story and not to ponder too long. Before the experiment started, participants went through four practice items and had the opportunity to ask any remaining questions.

In the experiment, first the trial number appeared on the screen for 1500 ms, accompanied by a 500 ms beep. Next, a fixation cross was shown for 600 ms, after which the first picture appeared. Immediately with the first picture, the first two context sentences were presented over the computer speakers. The second picture was presented 700 ms after termination of the second sentence, together with the word *Daarom* 'That's why'. Recording started at the same time. An 8 s pause followed, in which the second picture remained on the screen and the participant could complete the story. When the 8 s had elapsed, recording stopped and the next trial was started automatically. It took about 15 minutes to complete the experiment.

Data coding

After discarding the filler and practice items, the remaining (16 x 64 =) 1024 responses were scored for the type of referring expression used to refer to the target referent. The following codings were employed: NPs preceded by a definite article (*de man* 'the man') were coded as 'NP'; third person singular pronouns (*hij*, *ie/die* 'he', *zij*, *ze* 'she') were coded as 'pronoun'. In addition, reduced pronouns were also separately coded. However, since in contrast to the feminine reduced pronoun (*ze* 'she'), the masculine reduced pronoun (*ie/die* 'he') is a clitic with a restricted distribution, analyses were only performed on the feminine forms.

Only responses in which reference was made to the agent character in the second picture as a subject directly following *Daarom* and a finite verb were analyzed. We excluded 43 responses in which participants referred to the non-agent character, 2 cases in which reference was made to both characters at the same time, 5 cases in which the word *Daarom* 'That's why' was repeated, 3 cases in which the referring expression was not clear, and 2 cases in which there was no response. In all, 55 responses (5.4%) were excluded, equally spread over the conditions.

Design and statistical analyses

Crossing the two independent variables resulted in a 2 (target referent is + or – linguistically salient) x 2 (target referent is + or – visually salient) within-subjects and within-items design. The proportion of pronoun responses out of all responses and the proportion of reduced feminine pronoun responses out of all feminine pronoun responses were the dependent variables. We conducted two logit mixed model analyses (Jaeger, 2008): One over the proportion of pronoun responses, and one over the proportion of reduced feminine pronoun responses. In both cases, linguistic and visual salience of the target referent were included as fixed factors, and participants and items as random factors. One stimulus item was omitted from the analyses, because the overall proportion of pronouns in this item exceeded 2.5 standard deviations from the mean.

Results

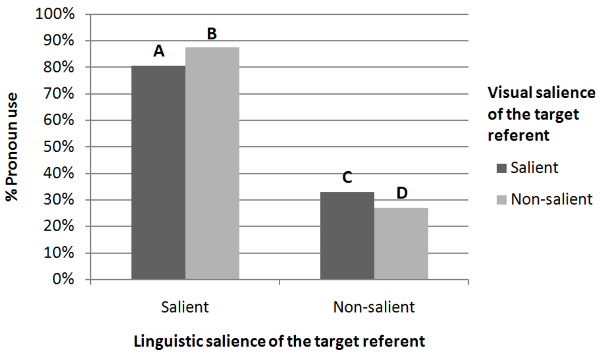

Figure 2: Percentage of pronoun references out of all references by linguistic and visual salience of the target referent (letters correspond to conditions in Fig. 1).

The results for the proportion of pronoun responses out of all responses are presented in Figure 2. We found a significant main effect of linguistic salience on pronoun use (β = 4.24, SE = 0.28, z = 15.27, p < .001): more pronouns were used when the target referent was linguistically salient. There was no main effect of visual salience on pronoun use (z < |1|). However, there was a significant interaction between linguistic and visual salience (β = -1.43, SE = 0.43, z = -3.36, p < .001), indicating that the effect of visual salience was different for linguistically salient referents than for linguistically non-salient referents. Planned comparisons showed that the effect of visual salience was significant both in the linguistically salient (β = -0.90, SE = 0.33, z = -2.71, p < .001) and in the linguistically non-salient (β = 0.58, SE = 0.27, z = 2.16, p = .03) condition. This means that when the target referent was linguistically salient, a *lower* visual salience led to more pronouns, while pronoun use increased with a *higher* visual salience when the target referent was linguistically non-salient. The inclusion of the random effects for participant and item ensures that the model controls for between-participants and between-items variance ($s^2 = 4.57$ and $s^2 = 0.11$, respectively).

Next, we investigated the proportion of reduced pronouns in a subset of the data including only the cases in which a feminine pronoun (*ze*, *zij* 'she') was used. The results are shown in Figure 3. We found a significant main effect of linguistic salience on the use of full versus reduced pronouns ($\beta = 2.72$, $SE = 0.64$, $z = 4.26$, $p < .001$): More reduced pronouns were used when the target referent was linguistically salient. There was a marginally significant effect of visual salience ($\beta = -1.10$, $SE = 0.56$, $z = -1.95$, $p = .05$), suggesting a tendency for more reduced pronouns when the target referent was visually non-salient. There was no significant interaction between linguistic and visual salience ($z < |1|$). The between-participants and between-items variances were $s^2 = 6.76$ and $s^2 = 0.60$, respectively.

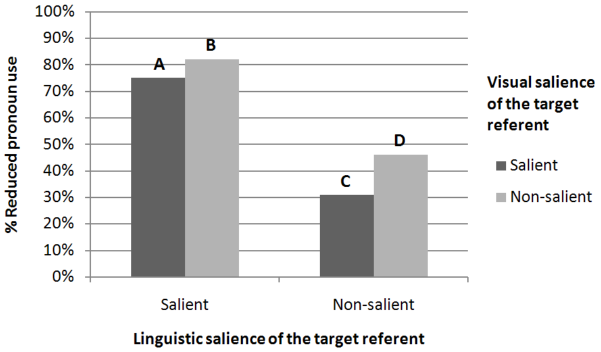

Figure 3: Percentage of reduced feminine pronoun references out of all feminine pronoun references by linguistic and visual salience of the target referent (letters correspond to conditions in Fig. 1).

Discussion

Our story completion experiment supports findings from other studies (e.g. Arnold, 1998; Gordon et al., 1993; Stevenson et al., 1994) that a referent's salience in the preceding linguistic context has an impact on the choice of referring expression: The likelihood of using a pronoun is higher when the referent is the subject of the directly preceding sentence than when it is not. In addition, the use of reduced pronouns as opposed to full pronouns also increases with a higher linguistic salience. While this contrasts with the finding by Kaiser and Trueswell (2004) that the full pronoun *zij* ('she') and the reduced pronoun *ze* are equally likely to refer to the subject of the preceding sentence, it is not incompatible with their suggestion that the use of full pronouns in Dutch is driven by contrast. It is possible that in the linguistically non-salient conditions, participants contrasted the target referent with the subject of the directly preceding sentence, which might explain the higher frequency of *zij* in these conditions.

More importantly, our results suggest that a referent's visual salience influences pronoun use. For linguistically non-salient referents, pronoun references increased with a higher visual salience. This is compatible with an accessibility-based account of reference. Entities occupying the foreground of a visual scene are more prominent in perception than entities in the background (Mazza et al., 2005). Therefore, visually salient referents have a representation in memory that is more activated and thus better accessible than that of less visually salient referents. As a result, expressions referring to visually salient entities tend to be more reduced. The fact that this effect was only found in the linguistically non-salient condition suggests that linguistic salience is still a more important factor in determining a referent's accessibility. This is in accordance with previous findings on the interaction between linguistic and visual information in reference production (Arnold & Griffin, 2007; Fukumura et al., 2010). When salience in the linguistic context is not decisive, visual properties of the referent may come into play in the choice of referring expression. In our experiment, the fact that a linguistically non-salient referent was still the subject of the first context sentence probably caused such a tempering effect on linguistic salience, as we intended. This might explain why a higher visual salience only led to more pronouns when the referent was linguistically non-salient.

For linguistically salient referents, however, pronoun references increased with a *lower* visual salience. In addition, the number of reduced pronouns tended to increase with visually non-salient referents. These findings are not predicted by an accessibility account. It might be the case that some other process is responsible for this effect. We tentatively propose that a clash between participants' expectancies as to which character the story is about and the actual story continuation may have caused processing difficulties, resulting in an increase of low-cost referring expressions. Recall that in the condition where the target referent was linguistically salient and visually non-salient (condition B in Figure 1), the other character was presented as the subject of the first context sentence ('Once upon a time there was…'). In addition, this character was already visually foregrounded in the first picture. This could have led participants to perceive this character as the protagonist. They could have seen the second context sentence as an aside, expecting the main story line to return to the protagonist. Indeed, protagonists have been found to remain accessible in a narrative, even after a topic shift (Anderson, Sanford, & Garrod, 1983). Analyzing the responses that were excluded because participants referred to the non-agent, however, did not reveal an effect of protagonisthood or visual salience on referent choice. Still, uncertainty in discourse understanding may occur when prominent characters are not involved in prominent events (Morrow, 1985). Thus, when the story continues with a visually non-salient character that was not the protagonist, more processing might be needed to integrate the unexpected event in the context and to formulate an utterance to describe that event. Consequently, speakers may turn to more economical expressions, such as pronouns, in case of processing difficulties (Almor, 1999; Ariel, 1990). Such an analysis might also explain our finding that reduced pronouns tend to be more frequent for visually non-salient referents, even in the linguistically non-salient conditions. Here, entities are apparently accessible enough in both linguistic contexts to be referred to with a pronoun. When they are involved in a visually non-salient event, their linguistic accessibility does not match the construction of the visual scene. This mismatch may lead to a larger effort in integrating the two modalities, resulting in more reduced forms.

A new study should address these issues by constructing the linguistic context in such a way that no expectations are raised about the upcoming event. For example, the characters could be introduced in a coordinated NP ('a boy and a girl had a conversation'). In addition, to investigate whether visual salience only increases accessibility when linguistic salience is indecisive, a condition should be included in which both characters are kept equally prominent in the story, such that the linguistic context does not impose a clear preference for a pronoun or a full NP.

In sum, the present study provides evidence that visually salient referents induce more pronoun references than visually non-salient referents, but only when they are not linguistically salient. This suggests that visual properties of referents affect accessibility, but can be overruled by linguistic properties. Future research should shed more light on the exact interplay between linguistic and visual information in the production of referring expressions.

Acknowledgements

We are grateful to Marjolijn Antheunis, Martijn Balsters, Ben Bergmans, Ed Boschman, Rein Cozijn, Constantijn Kaland, Ruud Koolen, Kitty Leuverink, Lisette Mol, Hanneke Schoormans, Mandy Visser, Hans Westerbeek and our other colleagues at TiCC for their help with the experiment and their useful comments.

References

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review, 106*(4), 748-765.

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *The Quarterly Journal of Experimental Psychology, 35*(3), 427-440.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (1998). *Reference form and discourse patterns.* PhD dissertation, Stanford University, Palo Alto, CA.

Arnold, J. E., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56*, 521-536.

Beun, R.-J., & Cremers, A. H. M. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition, 6*(1/2), 121-152.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54*, 592-609.

Clark, H. H., & Sengul, C. L. (1979). In search of referents for nouns and pronouns. *Memory and Cognition, 7*, 35-41.

Coco, M. I., & Keller, F. (2009). The impact of visual information on reference assignment in sentence production. *Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci)* (pp. 274-279). Amsterdam, The Netherlands.

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

Fukumura, K., Van Gompel, R., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *The Quarterly Journal of Experimental Psychology, 63*(9), 1700-1715.

Givón, T. (1983). *Topic Continuity in Discourse*. Amsterdam/Philadelphia: John Benjamins.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language, 57*(4), 544-569.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names and the centering of attention in discourse. *Cognitive Science, 7*(3), 311-347.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics, 21*(2), 203-225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of anaphoric expressions in discourse. *Language, 69*, 274-307.

Jaeger, T. F. (2008). Categorial data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Kaiser, E., & Trueswell, J. C. (2004). The referential properties of Dutch pronouns and demonstratives: Is salience enough? *Proceedings of the Sinn und Bedeutung 8, Arbeitspapier Nr. 1977*: FB Sprachwissenschaft, Universität Konstanz.

Kelleher, J., Costello, F., & Van Genabith, J. (2005). Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context. *Artificial Intelligence, 167*, 62-102.

Mazza, V., Turatto, M., & Umiltà, C. (2005). Foreground-background segmentation and attention: A change blindness study. *Psychological Research, 69*(3), 201-210.

Morrow, D. G. (1985). Prominent Characters and Events Organize Narrative Understanding. *Journal of Memory and Language, 24*(3), 304-319.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 94*, 473-592.