Language and Cognitive Processes

Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production

Jorrig Vogels [a] , Emiel Krahmer [a] & Alfons Maes [a]

[a] Tilburg center for Cognition and Communication (TiCC),
Tilburg University, Tilburg, the Netherlands

PLEASE SCROLL DOWN FOR ARTICLE

Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production

Jorrig Vogels, Emiel Krahmer, and Alfons Maes

Tilburg center for Cognition and Communication (TiCC), Tilburg University, Tilburg, the Netherlands

Salient entities are assumed to be more accessible in memory, which makes them more likely to be referred to first and to be referred to with an attenuated expression, such as a pronoun. It is less clear, however, how different types of salience interact in influencing referent accessibility. In this article, we address the question whether non-linguistic factors can affect accessibility in the presence of a linguistic context. We present two story completion experiments in which we investigated the effect of visual salience (foregrounding) in interaction with linguistic salience (subjecthood) of two story characters both on the choice of referent and on the choice of referring expression. In Experiment 1, linguistic salience was moderated by inducing a topic shift in the discourse context. In Experiment 2, contexts in which linguistic salience was unclear were compared to contexts in which one of the characters was highly linguistically salient. The results show that visual salience influences referent choice independently of linguistic salience, but that it does not have an effect on the choice of referring expression. This suggests that visual salience has an influence on the global interpretation of the scene, but does not directly affect the accessibility status of individual entities. This is compatible with a view of language production in which utterance planning is influenced by conceptual and discourse factors rather than by low-level perceptual factors.

Keywords: Visual salience; Reference; Referring expressions; Accessibility.

Reference is an important part of human communication. When we speak, we are constantly referring to objects or persons in our physical environment, to previous linguistic utterances, or to general knowledge about the world around us. A consequence of this is that our use of language is strongly interwoven with the context in which it is produced (Ariel, 1990). The ways in which context affects the production of references are not completely understood. Context may influence references in different parts of the production process. On the one hand, it may affect what people choose to refer to first, that is, how they choose a "starting point" for

their utterance (MacWhinney, 1977). On the other hand, when a referent has been established, context may influence what type of referring expression people use to refer to it, 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 is generally acknowledged that an important factor in guiding speakers both in what to refer to and in how to refer to it is salience. Salient entities are assumed to have a mental representation that is more activated, and hence are more readily available for the language production process (Levelt, 1989). Therefore, people tend to mention the most salient entity in the context first (e.g. Grosz, Joshi, & Weinstein, 1995; Osgood, 1971; Osgood & Bock, 1977; Stevenson, Crawley, & Kleinman, 1994). In addition, people tend to use more reduced expressions to refer to a salient entity, probably because highly accessible entities do not need an extensive description for identification (e.g. Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993).

Different factors, both linguistic and non-linguistic ones, have been identified as contributors to salience. An entity is more salient, for example, when it is a subject (Gordon, Grosz, & Gilliom, 1993; Grosz et al., 1995), a topic (Ariel, 1990; Givón, 1983; Gundel et al., 1993), the source of an event (Arnold, 2001; Stevenson et al., 1994), animate (Prat-Sala & Branigan, 2000 and many others), large (Flores d'Arcais, 1975), or more imageable (Bock & Warren, 1985). However, it is not clear how different types of salience interact in making a referent more or less accessible, and which types of salience are relevant for which part of the production process. Especially in research on the choice of referring expressions, the focus has been on the influence of salience in the linguistic context. The role of non-linguistic factors, such as visual salience, is much less clear. Given the fact that people often refer to things that are present both in their physical environment and in the discourse context, an important question is how visual salience affects reference in interaction with linguistic salience. In this article, we will explore to what degree visual salience is used in reference, and how it interacts with the linguistic context of an utterance. We will investigate both whether people are influenced by visual salience in choosing what to refer to first, and whether they are influenced by visual salience in their choice for a particular type of referring expression.

Research on the effect of the visual context on language production suggests that speakers use non-linguistic information in planning their utterances. For instance, visual attention has been found to affect syntactic structure. When people describe visual scenes in which visual attention is drawn to a particular figure, they are more likely to mention this figure first. When the figure is the patient of a transitive event, for example, people tend to use passives or predicates that take the perspective of the patient, such as "flee" instead of "chase" (Flores d'Arcais, 1975; Gleitman, January, Nappa, & Trueswell, 2007; Sridhar, 1988; Tomlin, 1997). These findings indicate that visually salient objects are more likely to be referred to first. It is not clear, however, how these effects of visual salience interact with other types of salience, such as the referent's discourse status. The studies cited above all made use of scene descriptions, in which no linguistic context was provided. Other studies have found that the salience of an entity in the linguistic context, for example, whether it constitutes given information or whether it was a subject in the preceding sentence, affects the choice of referent for first mention (e.g. Bock & Irwin, 1980; Ferreira & Yoshita, 2003; Prat-Sala & Branigan, 2000; Stevenson, 2002). Referent choice may, therefore, depend on a combination of linguistic and non-linguistic salience, or one type of salience might override the other (cf. Prat-Sala & Branigan, 2000 for the interaction between discourse salience and animacy).

There is also evidence that visual information influences the choice of referring expression. For example, people use more elaborate expressions, such as modified noun phrases instead of bare noun phrases, when multiple possible referents are present that are visually ambiguous (Brown-Schmidt & Tanenhaus, 2006; Ferreira, Slevc, & Rogers, 2005; Sedivy, 2003). They also use more reduced expressions when the referent is visually in focus (Beun & Cremers, 1998), and fewer reduced expressions when another possible referent is visually present, even when there is no ambiguity (Fukumura, Van Gompel, & Pickering, 2010). These findings suggest that the more salient a referent is in the visual context, the more attenuated the referring expression will be.

Most studies addressing the influence of visual information on the choice of referring expression were again scene description tasks that lacked a linguistic context. However, the choice of referring expression is generally assumed to be greatly influenced by the referent's saliency in the preceding discourse context. According to theories of reference (e.g. Ariel, 1990; Givón, 1983; Grosz et al., 1995), pronouns are appropriate when they refer to the subject or the topic of the previous sentence. Longer, more informative expressions are used when the referent is not a subject or a topic. Besides grammatical function and topicality, linguistic factors that have been found to affect the choice of referring expression include recency, givenness, frequency of mention, thematic roles, and syntactic position (e.g. Anderson, Garrod & Sanford, 1983; Ariel, 1990; Arnold, 1998; Clark & Sengul, 1979; Gernsbacher & Hargreaves, 1988; Givón, 1983; Gundel et al., 1993; Stevenson et al., 1994). Few studies have investigated effects of visual properties on the choice of referring expression when a linguistic context is present. This might be due to the assumption of some researchers (e.g. Ariel, 1998) that linguistic context is a far more important source of accessibility than the physical world. However, empirical evidence is typically lacking, the claim being mainly based on corpus studies of written texts (e.g. Ariel, 1998, 2001), a domain in which physical context generally does not play a large role. Recently, story completion experiments have been used to examine pronoun use in references to a character in the presence of both a visual and a linguistic context. In a study by Arnold and Griffin (2007), participants had to describe a picture following the presentation of another picture in combination with a context sentence. The stimuli came in three versions: (1) A competitor character was present both visually and linguistically; (2) no competitor was present; (3) a competitor was present both visually and linguistically in the first scene but not visually present in the second. Target referent and competitor were referentially unambiguous (i.e. of different genders). Arnold and Griffin found that participants used fewer pronouns to refer to the target character in condition (1) than in condition (2), suggesting that the presence of a competitor decreased the target referent's accessibility. However, no difference in pronoun use was found between conditions (1) and (3), suggesting that it was the linguistic presence of the competitor in the context sentence that affected salience rather than its visual presence during sentence production.

In a similar experiment, Fukumura et al. (2010) did find that visual context influenced the choice of referring expression. In their study, participants were presented with two consecutive pictures showing either one toy character (competitor not present condition) or two toy characters (competitor present condition). The first picture was combined with a written context sentence in which only the target character was mentioned (competitor not mentioned condition) or both characters were mentioned (competitor mentioned condition). In the second picture, the target character performed a simple action. Participants described this picture to a

confederate, who then acted out the description using the real toys. The results showed that participants used fewer pronouns to refer to the target referent when the competitor was visually present than when it was not visually present. In addition, the effect of visual context was larger in the condition where the competitor was linguistically present than in the condition where the competitor was not mentioned at all. These findings suggest that visual information affects the referent's accessibility and, therefore, the likelihood that it will be referred to with a pronoun. They also suggest that the visual context has a greater effect when the linguistic context is less compelling (i.e. when the competitor is also linguistically present).

Still, in both Arnold and Griffin (2007) and Fukumura et al. (2010), as well as other studies investigating the role of visual information on the production of referring expressions, visual information was manipulated differently than in the studies on referent choice cited above. The latter manipulated visual salience either by varying the objects' intrinsic perceptual properties, such as size (Flores d'Arcais, 1975; Osgood, 1971; Sridhar, 1988), or by making use of (implicit) attentional cues, such as a black square presented very shortly in the same position as where the target figure will appear immediately afterwards (Gleitman et al., 2007; Tomlin, 1997). The studies investigating choice of referring expression, however, varied the number of possible referents in the visual context, but not the intrinsic salience of the referents. With more than one possible referent, multiple entities have to be kept in memory at the same time. Because attention is spread over more than one entity, the individual activation of the entities may be reduced in the speaker's memory (cf. Arnold & Griffin, 2007). The influence of visual information on the choice of referring expression may thus be an effect of competition between possible referents, rather than being an effect of salience. To determine how salience affects the choice of referring expression, as well as the choice of referent, properties of the referent itself should be taken into account. Linguistically, referents that are topics or subjects are more salient than referents that are not topics or subjects (e.g. Arnold, 1998). Visually, properties such as size, centrality, color, foregrounding, orientation, intensity, and visual complexity have been identified as important cues to salience (e.g. Clark & Chase, 1972; Coco & Keller, 2009; Henderson & Ferreira, 2004; Kelleher, Costello, & Van Genabith, 2005; Mazza, Turatto, & Umiltà, 2005; Parkhurst, Law, & Niebur, 2002). The more prominent these perceptual properties, the more attention they will receive by the visual system (Desimone & Duncan, 1995; Parkhurst et al., 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 referent and the choice of referring expression in interaction with linguistic context.

Another question is at which stage of the production process effects of visual salience take place. According to Griffin and Bock (2000), order of mention (e.g. agent before patient) is influenced by a global apprehension of the event (e.g. what the source and the goal of the action are), but not by the visual salience of individual elements in the scene. Thus, some higher-level conceptual knowledge is required before linguistic processing can take place. For example, a mental model of the situation may be created, including world knowledge and inferences, which guides the formulation of utterances (Anderson et al., 1983; Johnson-Laird, 1983; Morrow, 1985; Sanford & Garrod, 1981). Low-level perceptual factors will not directly affect utterance planning (Bock, Irwin, & Davidson, 2004; Griffin & Bock, 2000). By contrast, Gleitman et al. (2007) found that even in events inducing a conjoined NP subject, in which there is no clear instigator or recipient of the action, visual attention influenced order of mention. They took this as evidence that attention directly affects the activation of lemma

representations, without the need for first apprehending the gist of the scene. If visual attention indeed leads to a higher accessibility of lemma representations, which subsequently affects order of mention in an incremental manner (e.g. Kempen & Hoenkamp, 1987), one would expect that it also affects the choice of referring expression, since this requires access to the corresponding lemma (Levelt, 1989). If, on the other hand, visual attention only influences the global representation of the scene, it may still affect the choice of referent by determining from which perspective the scene is viewed. For example, the same event may be seen as a ''giving'' event or as a ''receiving'' event, depending on which entity is taken as the figure and which as the ground (Gleitman et al., 2007). This may influence which entity is mentioned first. However, visual attention does not necessarily affect the choice of referring expression, since this is crucially dependent on the accessibility of mental representations of the individual entities, and not on the representation of the scene as a whole.

We hypothesise that if referent accessibility is dependent on a combination of linguistic and non-linguistic factors, a referent's visual salience will interact with its linguistic salience in determining the accessibility of its representation in memory. In this case, an entity that is salient in the visual context may activate its corresponding mental representation(s) more than other entities, in the same way as a linguistically salient entity does. This may guide the speaker in choosing what to mention first, as well as in the choice of how to encode it in a referring expression. Thus, visually salient referents are expected to be chosen more often as first referents and to be more often referred to with reduced expressions than visually non-salient entities. In addition, if linguistic information is more important in determining accessibility than visual information, as suggested by previous studies, any effects of visual saliency should at least be expected in contexts where linguistic saliency is less clear.

Alternatively, linguistic and visual salience may affect referential choices in different ways. Visual salience may influence the global representation of the narrative, for example, which are the main characters, but not directly affect the accessibility of individual entities, as suggested by Griffin and Bock (2000). Under this hypothesis, no effect of visual salience on choice of referring expression is predicted, while there may still be an effect on referent choice via figure-ground assignment or protagonisthood.

To investigate these alternative hypotheses, we conducted two story completion experiments in Dutch, in which we manipulated visual salience as a property of the characters in the visual scene themselves, and in which linguistic salience in two context sentences was also varied. In the experiments, two characters were always present simultaneously, both visually and linguistically, but one character was made visually and/or linguistically salient relative to the other. To make the stories naturalistic, the visual scenes used were photographs of real people. We investigated the effects of a referent's visual salience on the likelihood that it is mentioned first and on the use of pronouns versus full noun phrases. In Experiment 1, linguistic saliency was moderated by inducing a topic shift between the two context sentences. In Experiment 2, linguistic saliency was moderated by employing a noun phrase conjunction in the first context sentence and a sentence conjunction in the second context sentence. This condition was compared to two control conditions in which the linguistic saliency manipulation was very strong. In addition, differences in visual saliency were emphasised more than in Experiment 1. In both experiments, we found evidence for an effect of visual salience on referent choice, but not on choice of referring expression, suggesting that visual salience affects reference on a different level of processing than linguistic salience.
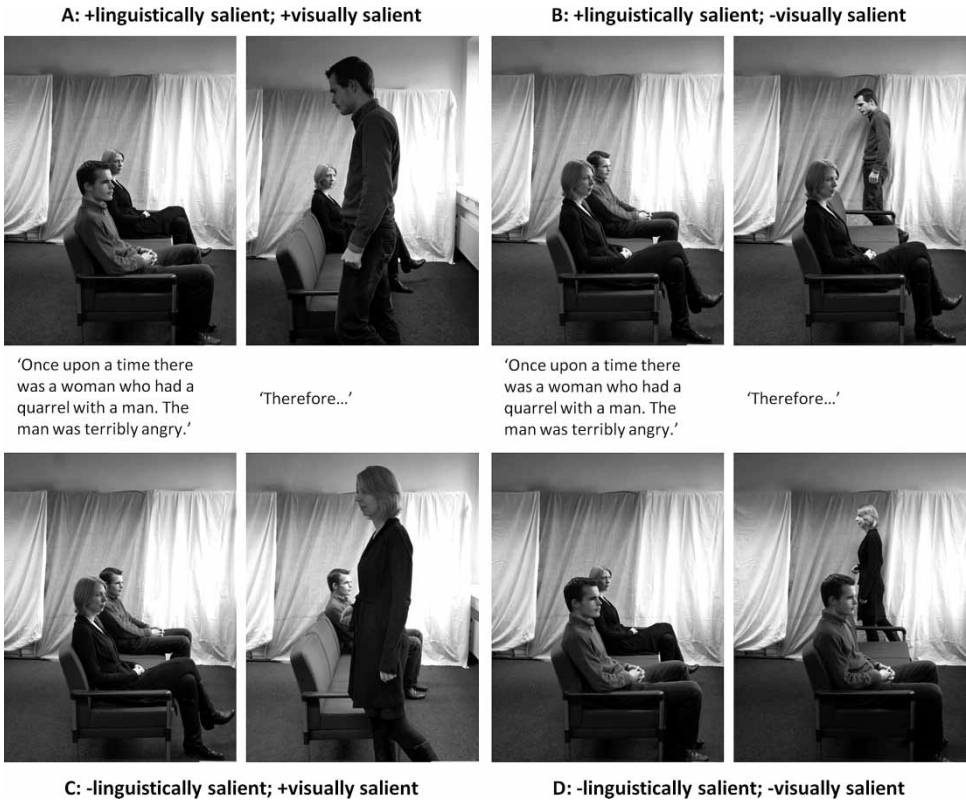
EXPERIMENT 1

Method

Participants

Sixty-four undergraduate students (17 male, 47 female; aged 18–43; mean age 21.7) from Tilburg University participated in the experiment for course credit. All were native speakers of Dutch and had normal or corrected to normal vision.

Materials

Twelve short stories were created that served as the experimental items. Each story consisted of two pictures, showing a male and a female character in a certain situation, accompanied by two sentences and the onset of a third sentence. The pictures and the sentences formed the context for the third sentence, which had to be completed by the participants. Sentences 1 and 2 accompanied the first picture of a pair, while the onset of the third sentence was aligned with the onset of the second picture. In the second picture, one character always performed an action (henceforth "agent character"). The other character did not move with respect to the first picture. Which character performed the action was varied across items. In addition, the visual and the linguistic salience of the characters were manipulated. The four different picture pairs of an experimental item with the accompanying context sentences are exemplified in Figure 1.

Linguistic salience was manipulated by making one of the characters the subject of the first context sentence and the other one the subject of the second context sentence. The subject of the second sentence was considered to be linguistically salient, because it was the subject of the sentence directly preceding the sentence that had to be completed, and the most recent possible antecedent for a referring expression. This is in line with theories of reference, such as centering (Grosz et al., 1995), in which the subject of the previous utterance is considered the most salient entity. The subject of the first sentence was considered to be linguistically less salient, because referential distance was longer (e.g. Ariel, 1990). Thus, one character was the subject of the second sentence (linguistically salient) and the other the subject of the first sentence (linguistically non-salient). The topic shift between the first and the second sentence was included to ensure that neither character became so linguistically salient that any effects of visual salience would be overruled.

The context sentences all had the same structure, which is illustrated in Table 1. 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 (*een vrouw* "a woman") in half of the cases and to the male character in the other half (*een man* "a man"). The subject was modified by a relative clause describing the situation (e.g. *die ruzie had* "who had a quarrel"), always followed by a prepositional phrase introducing the other character (e.g. *met een man* "with a man"). Subsequently, this character became the subject of the second sentence, which described a physical or emotional state (e.g. *De man was verschrikkelijk boos* "The man was terribly angry"). Across items, the predicates in the first two sentences were varied, as well as the gender of the subject and PP constituents (i.e. the subject of sentence 1 referred to the female character and both the character in the PP and the subject of sentence 2 referred to the male character, or vice versa). The onset of the third sentence always consisted of the word *Daarom* "Therefore". 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

A: +linguistically salient; +visually salient   B: +linguistically salient; -visually salient

'Once upon a time there
was a woman who had a
quarrel with a man. The
man was terribly angry.'   'Therefore...'

'Once upon a time there
was a woman who had a
quarrel with a man. The
man was terribly angry.'   'Therefore...'

C: -linguistically salient; +visually salient   D: -linguistically salient; -visually salient

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

constituent of interest. All context sentences were recorded by a female native speaker of Dutch. The sentences were pronounced with a neutral statement intonation, and with no stress accents on the noun phrases mentioning the characters (e.g. main stress in the second context sentence was always on the state describing adjective).

To test the consistency of the linguistic materials, we conducted a pretest in which participants provided written completions of the sentence combinations (without the pictures). The results showed no biases in particular items for the use of certain types of referring expression (One Way ANOVA with multiple comparisons: All p

TABLE 1
Template for the context sentences in Experiment 1

Sentence 1	Sentence 2	Onset of sentence 3
"Er was eens {een vrouw, een man} die {PREDICATE} met {een man, een vrouw}"	"{De man, De vrouw} {was, had} {STATE}"	"Daarom..."
"Once upon a time there was {a woman, a man} who {PREDICATE} with {a man, a woman}"	"{The man, The woman} {was, had} {STATE}"	"Therefore..."

Note: Alternatives for constituents that vary across items are between curly brackets.

values >.1). A second pretest in which participants had to choose between either a pronoun or an NP for a fixed referent also showed no significant differences between the items (all *p* values >.5).

Visual salience in the pictures was manipulated by having one of the characters appear on the foreground and in a central position in the picture, while the other one appeared in the background in a more peripheral position. In most cases the foregrounded character also partly occluded the backgrounded character. Varying both the visual salience of the two characters and the agent character in the second picture resulted in four versions of each picture pair (either the male or the female character was visually salient; either the male or the female character was the agent; see Figure 1). The pictures were photographs taken with a digital camera. Two couples posed for all pictures, which were all taken in the same room, with a neutral background. A statistical test of the size of both characters (size in pixels, measured from head to foot) confirmed that across pictures, the characters that were intended to be visually salient were depicted significantly larger than the characters that were intended to be non-salient ($t_{142} = 19.375$; $p < .001$). To avoid any effects of the left-to-right orientation of the characters in the pictures, mirror versions were created for all picture pairs (not shown in Figure 1). In all, each story had eight different versions.

In the first picture of each story, both characters were in a neutral position (e.g. sitting next to each other on the couch). In the second picture, either the male character performed a simple action (e.g. walking away), or the female character performed the same action. Care was taken that the third sentence could be finished in both versions of the second picture, that is, when the man was the agent and when the woman was the agent. Therefore, the action depicted in the second picture had to be compatible with the preceding context in both versions. For example, the action of walking away in reaction to the man being angry can be performed by both characters, since this is a plausible reaction for people both when they are angry themselves and when they are faced with anger from someone else.

In addition to the 12 experimental items, 16 filler items and 3 practice items were constructed. The fillers were identical to the experimental items, except for the fact that more variation was included in the situations: Five items had one male and one female character, in five items there was only one character and in another six items there were two characters of the same gender. In addition, the characters were not always referred to as "a man" and "a woman", but they also sometimes had roles like "a teacher" or "a salesman". There was only one version of each filler and practice item. The items were distributed over four lists according to a Latin Square design, such that each list contained one condition of a stimulus item. On each list, half of the items were in the mirrored version, and another four lists were created in which the other half of the items were mirrored, thus resulting in a total of eight lists. For each list, items were quasi-randomised, with the filler items in a fixed position and two experimental items never occurring in consecutive slots.

Procedure

Participants sat in a low noise cabin behind a computer screen. They had a keyboard at their disposal, which was only used to start the experiment. Between the keyboard and the computer screen was a microphone. The experiment was assembled and run with the E-Prime 2.0 software program. Participants were instructed to complete each story initiated by the two context sentences in such a way that it would fit with the situation shown in the second picture. They were not otherwise instructed

about the content of their responses: Participants were free to complete the stories in any way they liked, with the only restriction that their first sentence had to be connected to the word *Daarom* "Therefore". 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. The first three trials were practice trials, after which participants had the opportunity to ask any remaining questions.

Pressing the space bar started 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 sentences of the story were presented over the computer speakers. The second picture was presented 700 ms after termination of the second sentence, together with the word *Daarom*. Recording started at the same time. As soon as the word *Daarom* had sounded, there was an 8 s pause in which the second picture remained on the screen and the participant could complete the sentence. When the 8 s had elapsed, recording stopped and the next trial was started. It took about 10 minutes to complete the experiment.

Data coding

After discarding the filler and practice items, the remaining $(12 \times 64 =)$ 768 recordings of the participants' story completions were scored for two variables: (1) Which character participants referred to as the subject of the sentence that started with *Daarom* "Therefore" and (2) which type of referring expression participants used for this referent. For the choice of referent, we coded references to the agent character as "agent" and references to the non-agent character as "non-agent". For the type of referring expression, the following codings were employed: NPs preceded by a definite article (*de man* "the man"), a demonstrative (*die vrouw* "that woman") or an adjective (*de boze man* "the angry man") were coded as "NP"; full pronouns (*hij* "he", *zij* "she") and reduced pronouns (*ie, die* "he", *ze* "she") were coded as "pronoun". Other types of referring expression were not attested.

All first references after the word *Daarom* "Therefore" occurring as the subject of the completion sentence were scored. All other references were ignored. We excluded 23 responses from the data set in which reference was made to both characters at the same time, and 18 in which neither character was mentioned as the subject. In addition, we removed 10 cases in which the word *Daarom* was repeated, 5 cases in which the referring expression used was not clear, 4 cases in which the response suggested a misinterpretation of the story with respect to the pictures, and 2 cases in which the recording did not contain any speech. In all, 62 trials (8%) were excluded from the data set, resulting in 706 useful trials for the analysis.

Design and statistical analyses

Crossing the two independent variables resulted in a 2 (agent character is linguistically salient or linguistically non-salient) $\times 2$ (agent character is visually salient or visually non-salient) within-subjects and within-items design. The dependent variables were the proportion of subject references to the agent character out of all references and the proportion of pronoun references to the character mentioned as the subject out of all subject references to that character. For the latter variable, references were analysed separately for each character type (agent or non-agent). However, because there were too few data points for the non-agent references to perform

statistical analyses, only the results for the references to the agent character will be reported. We conducted two logit mixed model analyses (Jaeger, 2008), using the lme4 package in the R software program (www.r-project.org): One over the log odds of a subject reference to the agent character, and one over the log odds of a pronoun reference to the agent character. In both cases, linguistic and visual salience of the agent character were included as fixed factors, and participants and items as random factors. The fixed factors were centered, to reduce collinearity between predictors. Random intercepts and random slopes for participants and items were included to account for between-subject and between-item variation. Starting with a model with a full random effect structure, we used model comparisons to determine whether the inclusion of a random slope was justified by the data. Random slopes that did not contribute to the fit of the model according to a likelihood ratio test were removed. Only the final models will be reported.

Results

Choice of referent

In the majority of the cases (624, 88%), participants referred to the agent character as the subject of their response. Still, there were 82 responses (12%) in which participants referred to the non-agent character (i.e. the character that did not move with respect to the first picture) in subject position. There was no difference between the mirrored and the unmirrored versions of the pictures, that is, participants were not more likely to refer to the agent character when it was on the left than when it was on the right, $\chi^2(1) < 1$, $p = .98$.

Figure 2 shows the proportion of subject references to the agent character as a function of its linguistic and visual salience. The final logit mixed model is summarised in Table 2.

Firstly, we found an effect of linguistic salience on referent choice: Surprisingly, subject references to the agent character were more frequent (95.5%) when this
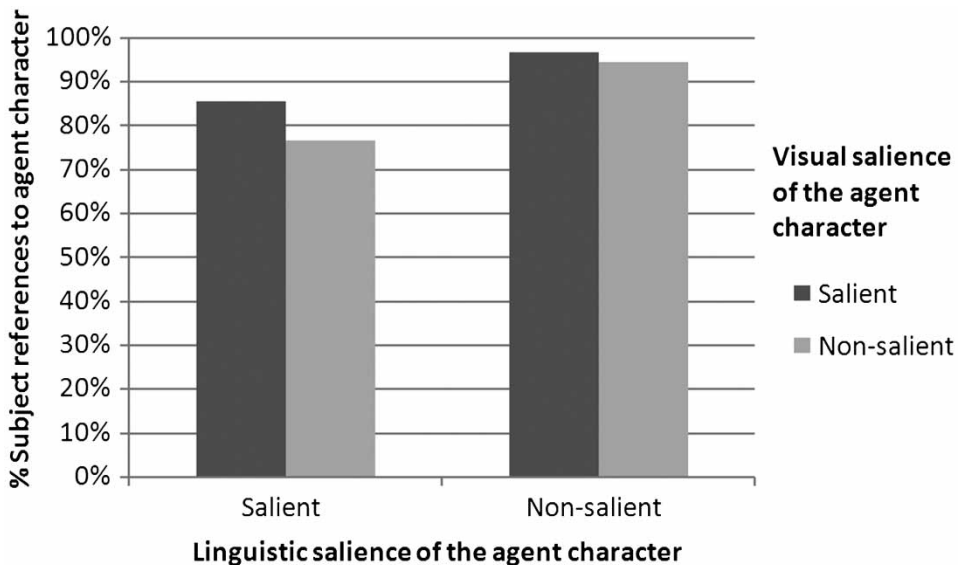
Figure 2. Percentage of subject references to the agent character out of all references in Experiment 1, plotted against its linguistic and visual salience.

TABLE 2
Logit mixed model for referent choice in Experiment 1

Random effects	s^2			
Subjects				
Intercept	1.45			
Ling. sal.	4.28			
Items				
Intercept	0.57			
Ling. sal.	9.49			
Fixed effects	β	SE	z	p
Intercept	3.81	0.39	9.85	< .001
Ling. sal.	− 2.28	1.06	− 2.15	< .05
Vis. sal.	0.82	0.41	1.99	< .05
Ling. sal. × Vis. sal.	0.15	0.81	0.19	.85

Notes: Ling. sal. = Linguistic salience; Vis. sal. = Visual salience.

character was *not* linguistically salient (i.e. when it was the subject of the first context sentence), than when it was linguistically salient (i.e. when it was the subject of the second context sentence) (81%). This difference was significant (see Table 2). The negative coefficient in Table 2 means that the predicted odds for a subject reference to a linguistically salient agent character are lower than the odds for a subject reference to a linguistically non-salient agent character.

Secondly, we found an effect of visual salience on referent choice: Subject references to the agent character were more frequent (91.1%) when this character was visually salient, than when it was visually non-salient (85.5%). This difference was significant (see Table 2). We found no significant interaction between linguistic and visual salience. By-subjects and by-items random slopes for linguistic salience were included, as they significantly improved the model's fit, $\chi^2(2) = 8.00$, $p < .05$ and $\chi^2(2) = 42.33$, $p < .001$, respectively. This indicates that participants and items varied in the degree to which they were affected by linguistic salience.

Choice of referring expression

For all subject references to the agent character ($n = 624$), we determined whether participants used a pronoun or a full noun phrase. The results for the effects of linguistic and visual salience on pronoun use are presented in Figure 3. The final logit mixed model is summarised in Table 3.

We found a main effect of linguistic salience: Pronouns were used more frequently when the agent character was linguistically salient (73%) than when it was linguistically non-salient (38%). This difference was significant (see Table 3). We found no main effect of visual salience (57% vs. 54% pronoun use), and no interaction between linguistic salience and visual salience. By-subjects and by-items random slopes were not included, as they did not improve the model's fit.

DISCUSSION

Experiment 1 showed effects of visual salience on the choice of referent, but not on the choice of referring expression. Participants used the visual salience of an entity to
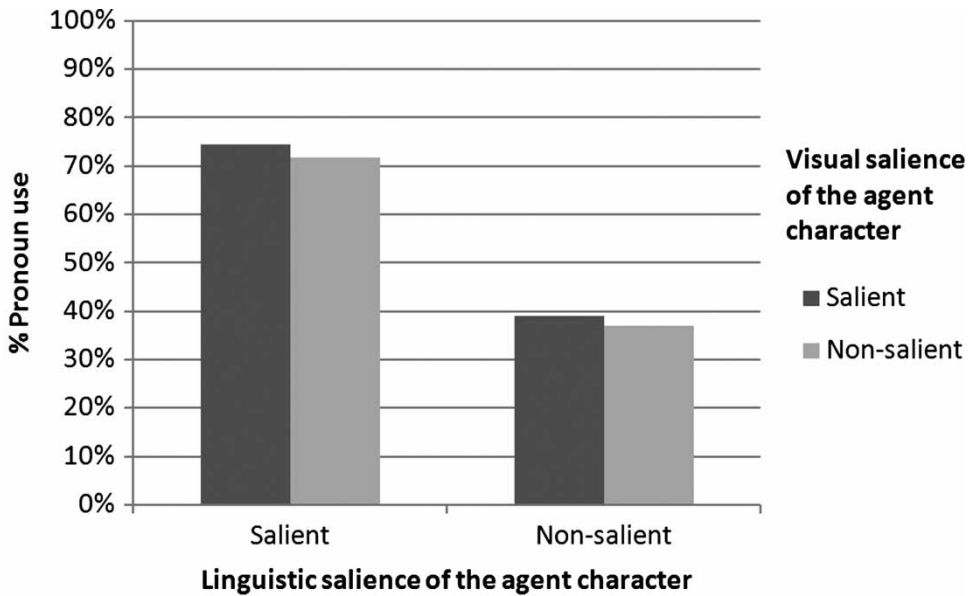
Figure 3. Percentage of pronoun references to the agent character out of all references (pronoun and full NP) to the agent character in Experiment 1, plotted against its linguistic and visual salience.

determine whether they would refer to it first (as the subject), but not to determine whether they would refer to it with a pronoun or a full NP. This suggests that visual salience does not affect the lemma representations associated with the individual characters in the scene, but is used for the global interpretation of the story, which subsequently influences figure-ground assignment or protagonisthood. In addition, there were no interactions with linguistic salience, indicating that the effects of visual salience were not influenced by whether the referent was the subject of the first or the second context sentence. This suggests that visual salience and linguistic salience affect reference production independently.

There may be two problems with the findings of Experiment 1. Firstly, the effect of linguistic salience on the choice of referent was not as predicted. The subject of the second context sentence, being the most recent subject, was expected to be more

TABLE 3
Logit mixed model for choice of referring expression in Experiment 1

Random effects	s^2			
Subjects				
Intercept	2.31			
Items				
Intercept	0.70			
Fixed effects	β	*SE*	*z*	*p*
Intercept	0.36	0.33	1.09	.27
Ling. sal.	2.25	0.23	9.93	<.001
Vis. sal.	0.16	0.21	0.78	.43
Ling. sal. × Vis. sal.	0.08	0.42	0.18	.86

Notes: Ling. sal. = Linguistic salience; Vis. sal. = Visual salience.

salient and hence to induce more subject references in the participants' response. However, participants were more likely to refer to the subject of the first context sentence than to that of the second. An example is given in (2c), (2a) and (2b) being the preceding context sentences (the participant's utterance is presented in bold).

(2) a. *Er was eens een man die stond te kletsen met een vrouw.*
"Once upon a time there was a man who was chatting with a woman''.
b. *De vrouw was een beetje duizelig.*
'The woman was a bit dizzy''.
c. *Daarom* **zei hij**$_{NON\text{-}AGENT}$ **dat ze**$_{AGENT}$ **een stoel moest pakken zodat ze erop kon gaan zitten**.
"Therefore he$_{NON\text{-}AGENT}$ said that she$_{AGENT}$ should take a chair so that she could sit down on it''.

In (2c), the subject pronoun *hij* "he'' refers to the subject *een man* "a man'' in (2a), even though in (2b) the subject has shifted to *de vrouw* "the woman'', and even though the referent is not the agent. This illustrates participants' preference to continue the story with the subject of the first context sentence, rather than with that of the second. We propose that this preference results from an interpretation of the subject of the first context sentence as the discourse topic or the protagonist of the story. The protagonist is the character from whose perspective the story unfolds. The story teller takes the viewpoint of the protagonist, and expresses this by putting it in subject or topic position right at the beginning of the discourse (Morrow, 1985). In the present experiment, the interpretation of the subject of the first sentence as the protagonist might have been encouraged by the fact that it was introduced with the words *Er was eens* "Once upon a time there was''. This may have caused the participants to continue referring to this character as the subject of their utterances (cf. Anderson et al., 1983).

This effect of protagonist assignment on referent choice may have had consequences for the effect of visual salience. Visual salience may also be a factor that affects the chance that a character is seen as the main character (or the "figure'') of the story. If linguistic and visual salience are indeed additive in determining the protagonist and therefore the choice of referent, this protagonisthood effect should be largest when the two factors are congruent, that is, when the referent is either both the subject of the first context sentence and visually salient, or both the subject of the second context sentence and visually non-salient. Figure 2 indeed suggests that this is the case. In addition, although we did not find a significant interaction between linguistic and visual salience, Figure 2 suggests that the effect of visual salience tends to be larger when salience is otherwise unclear (e.g. the referent is an agent but not linguistically introduced as the protagonist). Nevertheless, to single out the effect of visual salience on referent choice, linguistic contexts that are more neutral with respect to protagonist assignment should be investigated.

A second problem is that the fact that we did not find effects of visual salience on the type of referring expression could have been due to the dominance of linguistic salience: Information from the linguistic context could have been already sufficient to choose an expression, obscuring any effects of visual salience. It might be the case that visual salience only affects the choice of referring expression when a referent's salience in the linguistic context is unclear. It also might be the case that our manipulation of visual salience in Experiment 1 was not strong enough, or too heterogeneous across stimulus items to be manifest in the story continuations. In addition, the stimulus items may have differed in the degree of coherence between the context sentences.

This might have resulted in an interpretation of the second picture as unconnected to the preceding context. Experiment 2 was set up to deal with these issues.

EXPERIMENT 2

Experiment 2 was similar to Experiment 1, but a number of adaptations were employed to clarify the precise interplay between linguistic and visual salience in the choice of referent and the choice of referring expression. Firstly, we constructed context sentences in which both characters were expected to be about equally salient. These contexts were compared to contexts in which one character was made highly salient with respect to the other. If visual salience affects the choice of referring expression when linguistic salience is indecisive, an effect of visual salience is predicted in the former contexts, but not in the latter. Secondly, some adjustments to both the visual and the linguistic context were made to reduce differences in salience across items, to increase coherence between the context sentences, and to boost effects of visual salience.

Method

Participants

Forty-eight students (14 male; 34 female; aged 18 to 52; mean age 22.2) from Tilburg University participated for course credit. They were all native speakers of Dutch and had normal or corrected to normal vision. None of them had participated in Experiment 1.

Materials

Twelve new short stories were created, similar to the stimuli in Experiment 1. As before, they involved two characters, whose visual and linguistic salience were manipulated. Visual salience had two levels (salient or non-salient), and linguistic salience had three levels (salient, non-salient, or undetermined). This resulted in six versions of each stimulus item. An example of an experimental item is shown in Figure 4.

In the linguistically salient conditions (condition A and B in Figure 4), the agent character was the subject of both preceding context sentences. In the linguistically non-salient conditions (condition C and D in Figure 4), it was only mentioned in a prepositional phrase in the first context sentence. In the undetermined conditions (condition E and F in Figure 4), the second context sentence was a coordinated sentence, of which the first member had the agent character as the subject. The other character was mentioned as the subject of the second member. In the first context sentence, both characters were mentioned as the subject in a coordinated NP. This structure was chosen because it was assumed on the basis of the centering framework (Grosz et al., 1995) that when both characters are subjects in the directly preceding linguistic context, this makes their accessibility status unclear.

The sentences further differed from those in Experiment 1 in the following ways (see also Figure 4): Firstly, the first context sentence no longer started with *Er was eens. . .* "Once upon a time there was. . .", to avoid a strong linguistic cue for protagonist assignment. Secondly, instead of being exclusively referred to as *man* "man" and *vrouw* "woman" in the experimental items, the characters were called *jongen* "boy" and *meisje* "girl" in half of the items, to increase variation in the

A: +linguistically salient; +visually salient

'A boy was arguing with a girl. The boy got really annoyed.'

'Therefore…'

B: +linguistically salient; -visually salient

'A boy was arguing with a girl. The boy got really annoyed.'

'Therefore…'

C: -linguistically salient; +visually salient

'A boy was arguing with a girl. The boy got really annoyed.'

'Therefore…'

D: -linguistically salient; -visually salient

'A boy was arguing with a girl. The boy got really annoyed.'

'Therefore…'

E: ?linguistically salient; +visually salient

'A girl and a boy were arguing. The girl was full of good will but the boy got really annoyed.'

'Therefore…'

F: ?linguistically salient; -visually salient

'A girl and a boy were arguing. The girl was full of good will but the boy got really annoyed.'

'Therefore…'

Figure 4. A stimulus item from Experiment 2 in six different conditions: (A) Agent character is both linguistically and visually salient; (B) agent character is linguistically but not visually salient; (C) agent character is visually but not linguistically salient; (D) agent character is neither linguistically nor visually salient; (E) agent character is visually salient and neither linguistically salient nor linguistically non-salient; (F) agent character is visually non-salient and neither linguistically salient nor linguistically non-salient. The corresponding context sentences are translations of the Dutch originals.

descriptions for the characters. In the second context sentence, the adjective describing a physical or emotional state of the character(s) was always stage-level, that is, it denoted a temporary, event-like property, such as *geïrriteerd* "irritated" or *moe* "tired" (Carlson, 1977). In contrast to individual-level predicates, which describe more or less permanent states (e.g. "arrogant" or "honorable"), stage-level adjectives are more likely to induce an episodic reading of the story, and make it less likely that the second picture will be described as a habitual or generic event. In the linguistically undetermined conditions (E and F), the second sentence described a physical or emotional state of both characters, for example, *Het meisje was vol goede wil maar de jongen raakte enorm gepikeerd* "The girl was full of good will but the boy got really annoyed". In this way, both characters had the same thematic role (Experiencer), which should prohibit any differences in salience arising from this factor (see e.g. Stevenson et al., 1994). Finally, to further emphasise the episodic nature of the stories, the finite verb in the second sentence was changed from static *zijn* "to be" to dynamic *worden* "to become" or *raken/krijgen* "to get" (e.g. *het meisje werd tamelijk moe* "the girl became pretty tired"). All sentences were recorded by the same speaker as in Experiment 1. Care was taken that in condition E and F the coordinated clauses in the second context sentence formed a single prosodic unit, such that they would not be interpreted as separate sentences.

Visual salience was manipulated in the same way as in Experiment 1: The agent character in the second picture was either foregrounded (visually salient) or backgrounded (visually non-salient). However, some additional photographic means were employed to emphasise the difference in visual salience. Firstly, a spotlight was put on the character in the foreground, making that person appear brighter than the visually non-salient character. Secondly, the character in the background was blurred a little by putting the camera's focus on the character in the front. Next, measures were taken to avoid large differences in the manipulation of visual salience between items. First of all, the positions of the two characters were kept constant across items, such that the distance between the salient and the non-salient character was always the same. Secondly, the action in the second picture always involved at least standing up from a chair, causing the agent character to be upright at the onset of the third sentence. Furthermore, the only pieces of furniture used were two chairs and an optional table, and photographing was done against a white screen. This was done to minimise distraction from the two characters caused by other objects. Finally, the actions performed in the second picture of a pair were made more similar in appearance, each being 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, either for oneself or for the other character), or movements, in particular leaving/ walking away (e.g. after either the agent or the other character got irritated).

In addition to the 12 experimental items, 16 filler items and 4 practice items were constructed. These were similar to the filler items of Experiment 1. The items were distributed over a total of six stimulus lists. As in Experiment 1, mirror versions of all pictures were created. Since the left-to-right orientation of the characters in the pictures had not shown an effect on referent choice or choice of referring expression in Experiment 1, mirroring was done between items, instead of presenting each item in both a mirrored and an unmirrored version across lists. Items were distributed quasi randomly over each list.

Procedure

The procedure was similar to that of Experiment 1. After the participant had read the instructions, a practice block was started consisting of four practice items. Before the real experiment started, participants had the opportunity to ask any remaining questions. The setup of the experiment was the same as in Experiment 1. It took participants about 10 minutes to complete the experiment.

Data coding

After discarding the filler and practice items, $(12 \times 48 =)$ 576 responses to the experimental items remained. These were scored for referent choice and type of referring expression. The same coding scheme as in Experiment 1 was used. We excluded 1 response in which reference was made to both characters at the same time, and 1 response in which neither character was mentioned. We also excluded 1 response in which the word *Daarom* "Therefore" was repeated, and another response of which the syntax did not match the V2 structure initiated by *Daarom*. Finally, we excluded 3 cases in which the referring expression was unclear or missing. In all, seven responses (1.2%) were excluded, equally spread over the conditions, leaving 569 responses for analysis.

Design and statistical analyses

Crossing the two independent variables resulted in a 3 (agent character is linguistically salient, agent character is linguistically non-salient or linguistic salience is undetermined) $\times 2$ (agent character is visually salient or visually non-salient) within-subjects and within-items design. The dependent variables were the proportion of subject references to the agent character out of all references, and the proportion of pronoun references to the character mentioned as the subject out of all subject references to that character. As in Experiment 1, only the results for references to the agent character will be reported for the latter variable. Statistical analyses were the same as in Experiment 1. Because linguistic salience was now a predictor with three levels, it was recoded into two binary factors using contrast (sum) coding. The first factor represents the difference between linguistically non-salient characters and characters whose linguistic salience is undetermined. The second factor represents the difference between linguistically salient characters and characters whose linguistic salience is undetermined.

Results

Choice of referent

In 544 cases (96%), participants referred to the agent character as the subject of their utterance. In 25 cases (4%), the subject referred to the non-agent character. The results for the proportion of subject references to the agent character as a function of its linguistic and visual salience are presented in Figure 5. The final logit mixed model is summarised in Table 4.

Firstly, we found no effects of linguistic salience on referent choice: Although subject references were slightly more frequent for linguistically salient (i.e. the subject of both context sentences) agent characters (98%), than for linguistically undetermined (i.e. both characters were subjects in both context sentences) agent characters (96%), and slightly less frequent for linguistically non-salient (i.e. only
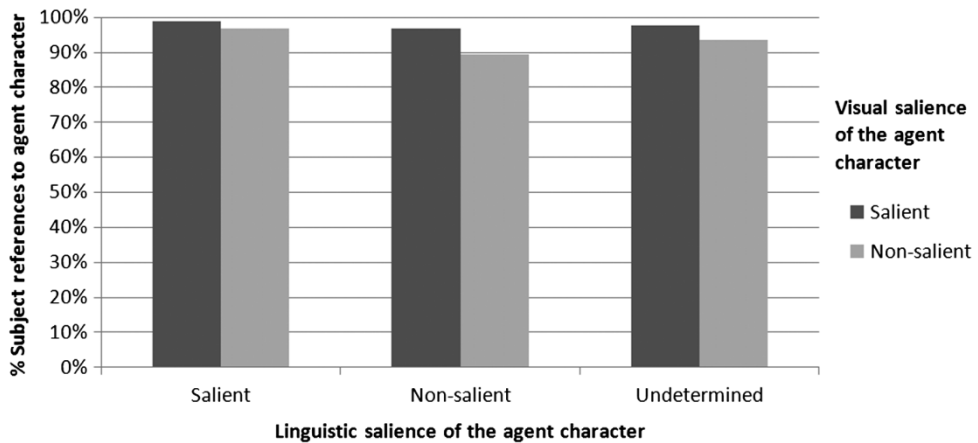
Figure 5. Percentage of subject references to the agent character out of all references in Experiment 2, plotted against its linguistic and visual salience.

present in a PP in the first sentence) agent characters (93%), these differences were not significant (see Table 4). A Tukey's test of multiple comparisons showed that the differences between all three levels were non-significant, $\beta = 0.94$, $SE = 3.65$, $z = 0.26$, $p = .96$ for the difference between linguistically salient and linguistically non-salient agent characters; $\beta = 0.16$, $SE = 2.41$, $z = 0.07$, $p = 1$ for the difference between linguistically undetermined and linguistically non-salient agent characters; and $\beta = -0.78$, $SE = 3.58$, $z = -0.22$, $p = .97$ for the difference between linguistically undetermined and linguistically salient agent characters. Using model comparisons, the overall main fixed effect of linguistic salience also turned out to be non-significant, $\chi^2(4) = 0.48$, $p = .98$.

TABLE 4
Logit mixed model for referent choice in Experiment 2

Random effects	s^2			
Subjects				
Intercept	11.77			
Ling. sal.: Non-sal. vs. Undet.	46.49			
Ling. sal.: Sal. vs. Undet.	74.74			
Items				
Intercept	1.85			
Fixed effects	β	SE	z	p
Intercept	9.33	1.47	6.33	< .001
Ling. sal.: Non-sal. vs. Undet.	−0.16	2.43	−0.07	.95
Ling. sal.: Sal. vs. Undet.	0.78	3.60	0.22	.83
Vis. sal.	2.25	0.84	2.67	< .01
Ling. sal.: Non-sal. vs. Undet. × Vis. sal.	0.23	1.77	0.13	.90
Ling. sal.: Sal. vs. Undet × Vis. sal.	−0.68	2.43	−0.28	.78

Notes: Ling. sal. = Linguistic salience; Vis. sal. = Visual salience; Sal. = Salient; Non-sal. = Non-salient; Undet. = Undetermined.

Secondly, we found an effect of visual salience on referent choice: References to the agent character were more frequent when it was visually salient (98%) than when it was visually non-salient (93%). This difference was significant (see Table 4). We found no significant interactions between linguistic and visual salience. By-subjects random slopes for linguistic salience were included, as they significantly improved the model's fit, $\chi^2(5) = 13.59$, $p < .05$, indicating that participants differed in the way they were affected by linguistic salience.

Choice of referring expression

For all subject references to the agent character ($n = 544$), we determined whether participants used a pronoun or a full noun phrase. The results for the effects of linguistic and visual salience on pronoun use are presented in Figure 6. The final logit mixed model is summarised in Table 5.

Firstly, we found an effect of linguistic salience: Pronouns were used more frequently (24%) when the linguistic salience of the agent character was undetermined than when the agent character was linguistically non-salient (11%), but less frequently than when it was linguistically salient (88%). The difference between linguistically non-salient and linguistically undetermined agent characters was significant, as was the difference between linguistically salient and linguistically undetermined agent characters, (see Table 5). A Tukey's test of multiple comparisons showed that the differences between all three levels were significant, $\beta = 16.57$, $SE = 2.35$, $z = 7.04$, $p = < .001$ for the difference between linguistically salient and linguistically non-salient agent characters; $\beta = 3.43$, $SE = 0.93$, $z = 3.70$, $p < .001$ for the difference between linguistically undetermined and linguistically non-salient agent characters; and $\beta = -13.14$, $SE = 2.27$, $z = -5.79$, $p < .001$ for the difference between linguistically undetermined and linguistically salient agent characters. Using model comparisons, the overall main fixed effect of linguistic salience was found to be significant, $\chi^2(4) = 408.27$, $p < .001$.

Secondly, we found only a marginally significant effect of visual salience (see Table 5), showing a small tendency for fewer pronouns for visually salient agent characters. There was also a marginally significant interaction between linguistic
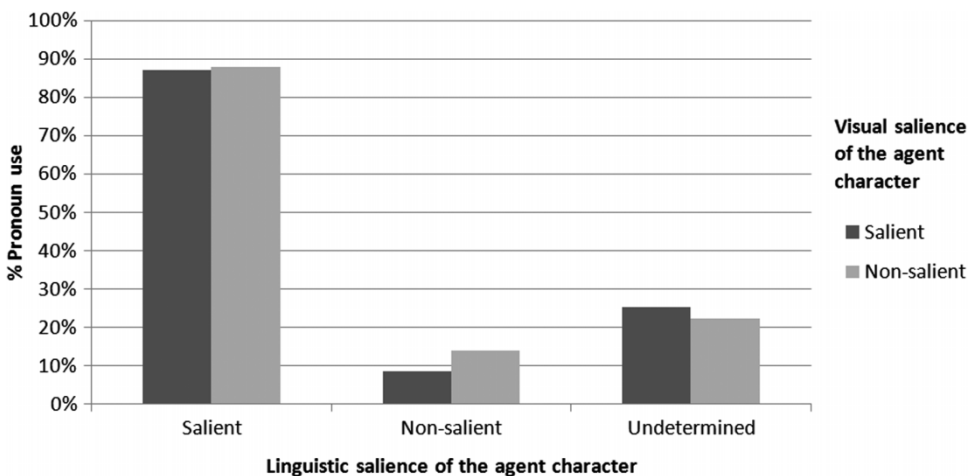
Figure 6. Percentage of pronoun references to the agent character out of all references (pronoun and full NP) to the agent character in Experiment 2, plotted against its linguistic and visual salience.

TABLE 5
Logit mixed model for choice of referring expression in Experiment 2

Random effects	s^2			
Subjects				
Intercept	50.16			
Vis. sal.	21.83			
Items				
Intercept	1.32			
Fixed effects	β	SE	z	p
Intercept	−2.60	1.22	−2.12	<.05
Ling. sal.: Non-sal. vs. Undet.	−3.46	0.92	−3.75	<.001
Ling. sal.: Sal. vs. Undet.	13.25	2.30	5.77	<.001
Vis. sal.	−1.70	1.01	−1.69	.09
Ling. sal.: Non-sal. vs. Undet. × Vis. sal.	−3.33	1.74	−1.91	.06
Ling. sal.: Sal. vs. Undet × Vis. sal.	3.01	2.18	1.39	.17

Notes: Ling. sal. = Linguistic salience; Vis. sal. = Visual salience; Sal. = Salient; Non-sal. = Non-salient; Undet. = Undetermined.

salience and visual salience (see Table 5), suggesting that the trend for fewer pronouns for visually salient referents only holds for linguistically non-salient referents. The difference in the effect of visual salience between linguistically salient and linguistically undetermined referents was non-significant. Recoding the predictor for linguistic salience revealed that the difference in the effect of visual salience between linguistically salient and linguistically non-salient referents was significant, $\beta = 6.49$, $SE = 2.37$, $z = 2.74$, $p < .01$: For linguistically salient referents, being visually salient results in more pronoun use than for linguistically non-salient referents. However, model comparisons showed that the overall interaction between linguistic and visual salience was only marginally significant, $\chi^2(2) = 4.68$, $p = .096$.

A by-subjects random slope for visual salience was included, as it significantly improved the model's fit, $\chi^2(2) = 9.89$, $p < .01$. This suggests that participants varied in the way they were affected by the manipulation of visual salience. Without the random slope, the main effect of visual salience and the interaction between linguistic and visual salience lose their (marginal) significance, which may suggest that these effects do not generalise to all participants.

DISCUSSION

Experiment 2 was conducted to investigate whether a character's visual salience influences referent choice and the choice of referring expression when the character's salience in the linguistic context remains unclear. In such a case, the information from the linguistic context might be insufficient for a speaker to experience a strong preference to continue the story with one of the characters, or to choose a particular type of referring expression. The salience of the character in the visual context might then become decisive in making these choices. In Experiment 2, we attempted to create such a situation by on the one hand including a condition in which both characters were salient (i.e. subjects) throughout the linguistic context, and on the other hand strengthening the manipulation of visual salience.

The results show that, as in Experiment 1, visual salience affects the choice of referent, and does so independently of linguistic salience. Conversely, linguistic salience strongly affects the choice of referring expression, as expected, but it does not influence the choice of referent, which is different from what we found in Experiment 1. This may be related to the difference in the manipulation of linguistic salience between the two experiments (i.e. whether or not one of the characters was introduced as the protagonist by the clause *Er was eens…* "Once upon a time there was…"). The results of Experiment 2 further reveal a small effect of visual salience on the choice of referring expression that we did not find in Experiment 1: Referents that are visually salient are slightly *less* likely to be referred to with a pronoun. The interaction effects suggest that this effect is mainly carried by references to linguistically non-salient referents (see also Figure 6). Thus, for linguistically non-salient referents, which are already unlikely to be referred to with attenuated expressions, being visually salient reduces the chances of a pronoun even more. One possible explanation for this is that visual salience matches with linguistic focus: In the linguistically non-salient conditions (C and D), the referent is linguistically in focus in the first context sentence. Focus is associated with high informational importance and new information. Visual salience might convey similar properties. Thus, the tendency to use full NPs to refer to a linguistically focused referent might be encouraged when this referent is visually salient.

However, the fact that the effects of visual salience disappear when the by-subjects random slope for visual salience is removed suggests that they may be due to deviant behaviour of some participants. Indeed, close inspection of the data revealed two participants that used full NPs only for linguistically non-salient referents that were visually salient. Without these two participants, the percentage of pronoun use in the linguistically non-salient conditions is about equal between the two visual conditions (around 9%). For linguistically salient referents and for referents whose linguistic salience is undetermined, visual salience does not seem to affect pronoun use, replicating our findings from Experiment 1. Thus, Experiment 2 shows that even when the referent's salience from the linguistic context is unclear, visually salient referents are not referred to with more attenuated expressions. This suggests that the lack of an effect of visual salience on the choice of referring expression in Experiment 1 is not due to the dominance of the linguistic context.

These results again suggest that the decision which character to mention first (as the subject) and the choice of referring expression are two processes that are affected by different kinds of information. Firstly, our finding that in the linguistically undetermined conditions (E and F) visual salience affects the choice of referent but does not affect the choice of referring expression suggests that visual salience adds to the overall interpretation of the scene by influencing which character is perceived as most important, but does not directly affect the accessibility of individual entities. This is compatible with the view that the order of mention of characters in a scene is affected by a global apprehension of the event (cf. Griffin & Bock, 2000).

Secondly, the lack of an effect of linguistic salience on referent choice suggests that in Experiment 2 the linguistic context did not imply a clear protagonist, because it did not explicitly introduce one of the characters as such (as opposed to *Er was eens…* "Once upon a time there was…" in Experiment 1). Thus, in Experiment 2, linguistic salience was solely defined in terms of subjecthood, which did not, as in Experiment 1, affect the choice of referent. This would also support the view that referent choice is affected more by global aspects of the narrative than by local saliency. Another explanation for the absence of a linguistic salience effect could be that it is due to the

large preference of participants to mention the agent character first. This might have overridden any effects of the linguistic context. However, it is not clear why this preference did not (completely) override the effect of visual salience in Experiment 2, or the effect of linguistic salience in Experiment 1, in which there was also a large preference to start the sentence with the agent character.

GENERAL DISCUSSION

Two story completion experiments showed that an entity's visual salience, as measured by its position in the foreground or in the background of the scene, influences the choice of referent, but not the choice of referring expression. Participants were more likely to start their utterance with the agent character when this character was visually salient, but they were not more likely to refer to this character with a pronoun. In addition, visual salience did not interact with linguistic salience, that is, whether the referent was the subject of the previous sentence. In contrast to visual salience, linguistic salience affected the choice of referring expression, but influenced referent choice only in Experiment 1, where participants were more likely to take the subject of the first context sentence as the protagonist.

The results of the present study follow up on earlier findings in the following ways. Firstly, the finding that the choice of referent is influenced by the perceptual properties of the characters supports findings from scene description tasks that showed that visual and attentional cues influence order of mention (Gleitman et al., 2007; Osgood, 1971; Osgood & Bock, 1977; Tomlin, 1997). Whereas such tasks typically elicited isolated sentences, we have shown that these effects also apply in narrative discourse contexts, in which the visual entities being mentioned are also available from the preceding linguistic context. This means that an entity's linguistic salience does not override effects of perceptual salience. In fact, only when one of the characters was linguistically marked as a protagonist (using the phrase *Er was eens...* "Once upon a time there was...", Experiment 1), the linguistic context had an influence on the choice of referent. This mirrors the effect of linguistic salience in Prat-Sala and Branigan (2000), who also found that the entity that had been introduced in an existential construction (*There was...*) was more likely to be referred to first. In line with Prat-Sala and Branigan's findings, our results support the view that it is a combination of linguistic and non-linguistic factors that determines order of mention.

Secondly, our results support earlier findings (e.g. Arnold, 1998; Gordon et al., 1993; Stevenson et al., 1994) that a referent's salience in the linguistic context has an impact on the choice of referring expression: The likelihood of choosing a pronoun over a full noun phrase is higher when the referent is the subject of the directly preceding sentence than when it is not mentioned in the previous sentence. More importantly, our results show, in contrast to Fukumura et al. (2010), that the choice of referring expression is not affected by the salience of the referent in the visual context, even when the referent's salience from the linguistic context is unclear. Our manipulation of visual salience was different from that by Fukumura and colleagues. While they varied the number of possible referents in the scene, our scenes always involved the same number of characters, but varied in the position of the characters relative to the observer (or more precisely, the camera). In visual perception, foregrounded objects have been found to be attended to more than objects in the background (Mazza et al., 2005). This might be a more sound manipulation of salience, since the one employed by Fukumura et al. also changes the number of items

that have to be kept in memory, affecting cognitive load and therefore accessibility (cf. Arnold & Griffin, 2007).

Our finding that visual salience affects the choice of referent but not the choice of referring expression supports the hypothesis that people use different types of information in choosing a subject referent and choosing a referring expression. The fact that visually foregrounded characters are more likely to be referred to as subjects indicates that the spatial location of the characters in the pictures is a salient feature that influences language production. However, this information is apparently not used in choosing a referring expression. Thus, our results suggest that the choice of referring expression should be dissociated from the choice of referent. While the choice of referring expression is likely to be dependent on the accessibility of mental representations (Ariel, 1990; Gundel et al., 1993), this does not automatically imply that the same information is also used to pick a referent for first mention. Rather, referent choice may be affected by protagonist or figure-ground assignment, which may be separate from the accessibility of the mental representations associated with individual entities (cf. Gleitman et al., 2007). In our experiments, people may have been more inclined to take the perspective of the visually foregrounded character. Hence, they may have interpreted it as the main character or figure in the discourse, which made it more likely that they referred to this character first. This account is consistent with the view that speakers primarily plan their utterances based on the global structure of an event, for example, concerning the relations between the entities in the scene, and less by taking into account the salience of individual elements (Bock et al., 2004; Griffin & Bock, 2000).

Similarly, in the linguistic context, the character introduced as the protagonist is more likely to be mentioned as the subject in the continuation, whereas the choice of referring expression is more dependent on the most recent subject (Gordon et al., 1993; Stevenson, 2002). While it has been found that protagonisthood may also affect the choice of referring expression (e.g. Anderson et al., 1983; Karmiloff-Smith, 1981), our results suggest that local discourse salience is a stronger factor here (cf. Van Vliet, 2008). This might be an indication that referent choice is primarily determined by global conceptual aspects of the narrative, such as which is the main character in the discourse, while the choice of referring expression seems to depend more on the model the speaker has of the preceding discourse, such as the referent's accessibility status in the immediate linguistic context.

The discrepancy that we find between the effects of linguistic and visual salience might originate in the way salience is defined. One could argue that salience in the visual context and salience in the linguistic context are in fact two different notions of salience. While linguistic salience may be seen as a property of an entity's representation in the discourse, visual salience is a property of the stimulus itself. Linguistic salience is often associated with topicality and predictability (e.g. Arnold, 2008, 2010; Givón, 1983): Salient referents are those referents that are likely to be mentioned again in the discourse, given their discourse status. Since predictable referents are more accessible, they are also more likely to be referred to with attenuated expressions. Visually salient entities, on the other hand, may attract attention, but it is unclear whether this necessarily increases the accessibility of their representations (cf. Arnold, 2010). It might even be conceivable that visual salience, at least in the way it was manipulated in the present study, is more associated with focus and unpredictability: It could be the case that visually foregrounding an entity marks it as important and having high news value. Similarly, visually backgrounding an entity might correspond to the linguistic notion of background, which contains given,

already established information (e.g. De Swart & De Hoop, 2000). We leave the discussion of the exact informational status of visual salience to further research. In addition, it remains to be seen how manipulating visual salience using other perceptual properties, such as a referent's inherent size, its color, dynamics, or affective properties, might affect reference.

Another possible explanation for the fact that we did not find that visually salient entities are referred to with more attenuated expressions is that visual salience might be too subtle to affect the choice between nominal and pronominal expressions. It could be the case that this division is too rough. Effects of visual salience might be only visible on a more fine-grained scale. To test this possibility, we investigated the use of reduced pronouns in Experiment 1 (Experiment 2 contained too few reduced pronouns to perform statistical analyses). Our experiments were conducted in Dutch, which has a distinction between full (*hij* "he", *zij* "she") and reduced (*ie/die* "he", *ze* "she") pronouns. Because the syntactic distribution of masculine reduced pronouns is more restricted than that of the feminine forms (they cannot occur sentence-initially), we only investigated the feminine reduced pronouns, which were compared to the feminine full pronouns. We found no significant effects of either linguistic or visual salience on the use of reduced pronouns, suggesting that salience does not play a large role at this level.

A further issue concerns the linguistically undetermined conditions in Experiment 2. These were created to make sure that effects of linguistic salience would not dominate those of visual salience. However, there was still a strong preference to use a full NP when referring to the agent character. This preference could have obscured effects of visual salience. It is possible that the preference to use a full NP is due to competition between the two characters. In the linguistically undetermined conditions, both characters were made subjects in the preceding context. It was predicted that they would be equally salient in the linguistic context, making the relative salience of the two characters unclear. Because the two characters are salient in the mind of the speaker, a large amount of attention may be allocated to both of them. However, if attention is a limited resource, the amount of attention captured by each entity should be lower than when all attention can be allocated to one salient character (Arnold & Griffin, 2007). Hence, the accessibility of each character might actually be lower, which results in more elaborate referring expressions, such as full NPs. It could also be the case that because the context sentences in our experiments were presented auditorily and the participants had to respond orally within a fixed time frame, participants had a hard time keeping track of the characters mentioned in the linguistic context. As a result, they might have taken the last mentioned character as the most salient, that is, of which activation was not yet decaying (e.g. Almor, 1999). To test this, we conducted another experiment with the same materials as Experiment 2, but with written rather than spoken context sentences. In addition, participants had to write their response down while they could still view the context sentences and could take all the time they liked. The results of this experiment were very similar to that of Experiment 2, with only slightly more pronouns in the linguistically undetermined condition. This indicates that the extensive use of full NPs in this condition is probably not due to a fast decay of memory activations.

Nevertheless, both in Experiment 2 and in the written experiment, the proportion of pronouns used to refer to the agent character in the linguistically undetermined condition was significantly higher than that in the linguistically non-salient condition, in which the agent character had only been mentioned in a PP adjunct, and significantly lower than that in the linguistically salient condition, in which the agent

character had been the only subject in the preceding two sentences. This indicates that the salience of the agent character in the linguistically undetermined condition was indeed less clear than in the other conditions.

CONCLUSION

In summary, the present study has shown that a character's visual salience in a narrative discourse guides speakers in which character to refer to first, as the subject of their utterance, but not in what kind of referring expression they use. These results suggest that visual salience does not affect the accessibility of individual entities. Rather, it influences the global interpretation of the situation. Our results are compatible with the view that utterance planning is not directly influenced by low-level perceptual properties, but that at least some higher-level conceptual information is required for the formulation of a message. Our results also suggest that the choice of referring expression is more dependent on the status of the speaker's current discourse model, and hence works on a more local level.

REFERENCES

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748–765. doi:10.1037/0033-295X.106.4.748

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.

Ariel, M. (1998). The linguistic status of the "here and now". *Cognitive Linguistics*, *9*(3), 189–237.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins.

Arnold, J. E. (1998). *Reference form and discourse patterns* (Doctoral dissertation, Stanford University, Palo Alto, CA, USA). Retrieved from http://www.unc.edu/~jarnold/papers/diss/fulldissertation.doc.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527. doi:10.1080/01690960801920099

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203. doi:10.1111/j.1749-818x.2010.00193.x

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. doi:10.1016/j.jml.2006.09.007

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

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 467–484. doi:10.1016/S0022-5371(80)90321-7

Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action. Eye movements and the visual world* (pp. 249–278). New York and Hove: Psychology Press.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67. doi:10.1016/0010-0277(85)90023-X

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. doi:10.1016/j.jml.2005.12.008

Carlson, G. N. (1977). A unified analysis of the English bare plural. *Linguistics and philosophy*, *1*, 413–456. doi: 10.1007/BF00353456

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. doi:10.1016/0010-0285(72)90019-9

Clark, H. H., & Sengul, C. L. (1979). In search of referents for nouns and pronouns. *Memory and Cognition*, *7*, 35–41. doi:10.3758/BF03196932

Coco, M. I., & Keller, F. (2009). The impact of visual information on reference assignment in sentence production. In N. A. Taatgen & H. van Rijn (Eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 274–279). Austin, TX: Cognitive Science Society.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

De Swart, H. E., & De Hoop, H. (2000). Topic and focus. In L. Cheng & R. Sybesma (Eds.), *The first glot international state-of-the-article book* (pp. 105–130). Berlin and New York: Mouton de Gruyter.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263–284. doi:10.1016/j.cognition.2004.09.002

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, *32*(6), 669–692. doi:10.1023/A:1026146332132

Flores d'Arcais, G. B. (1975). Some perceptual determinants of sentence construction. In G. B. Flores d'Arcais (Ed.), *Studies in perception: Festschrift for Fabio Metelli* (pp. 344–373). Milan, Italy: Martello-Guinti.

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.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*(6), 699–717. doi:10.1016/0749-596X(88)90016-2

Givón, T. (1983). *Topic continuity in discourse*. Amsterdam, the Netherlands and Philadelphia, PA: 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. doi:10.1016/j.jml.2007.01.007

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.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. doi:10.1111/1467-9280.00255

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.

Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The interface of language, vision, and action. Eye movements and the visual world*. New York and Hove: Psychology Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Karmiloff-Smith, A. (1981). The grammatical marking of thematic structure in the development of language production. In W. Deutsch (Ed.), *The child's construction of language* (pp. 121–148). London: Academic Press.

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. doi:10.1016/j.artint.2005.04.008

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.

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

MacWhinney, B. (1977). Starting points. *Language*, *53*(1), 152–168.

Mazza, V., Turatto, M., & Umiltà, C. (2005). Foreground-background segmentation and attention: A change blindness study. *Psychological Research*, *69*(3), 201–210. doi:10.1007/s00426-004-0174-9

Morrow, D. G. (1985). Prominent characters and events organize narrative understanding. *Journal of Memory and Language*, *24*(3), 304–319. doi:10.1016/0749-596X(85)90030-0

Osgood, C. E. (1971). Where do sentences come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 497–529). London: Cambridge University Press.

Osgood, C., & Bock, J. K. (1977). Salience and sentencing: Some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 89–145). Hillsdale, NJ: Erlbaum.

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. doi:10.1016/S0042-6989(01)00250-4

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168–182.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Explorations of comprehension beyond the sentence*. Chichester and New York: Wiley.

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. doi:10.1023/A:1021928914454

Sridhar, S. N. (1988). *Cognition and sentence production: A cross-linguistic study*. New York: Springer-Verlag.

Stevenson, R. (2002). The role of salience in the production of referring expressions. In K. V. Deemter & R. Kibble (Eds.), *Information sharing: Reference and presupposition in language generation and interpretation* (pp. 167–192). Stanford, CA: CSLI Publications.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *94*, 473–592. doi:10.1080/01690969408402130

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). New York: Cambridge University Press.

Van Vliet, S. (2008). *Proper nouns and pronouns. The production of referential expressions in narrative discourse*. Utrecht: LOT