

This is a preprint of the manuscript entitled “Conceptual match as a determinant of reference reuse in dialogue”, to be published in the Journal of Experimental Psychology: Learning, Memory, and Cognition. Small differences between this version and the final version of the manuscript might be found.

Running head: Conceptual match effects in dialogue

Conceptual match as a determinant of reference reuse in dialogue

Dominique KNUTSEN

Department of Psychology, University of Essex, UK

Ludovic LE BIGOT

University of Poitiers & CNRS (CeRCA, UMR 7295), France

Corresponding author

Dominique Knutsen

University of Essex

Wivenhoe Park

Colchester CO4 3SQ

United Kingdom

Email address: dknutsen@essex.ac.uk

Abstract

As speakers interact, they add references to their common ground, which they can then reuse to facilitate listener comprehension. However, all references are not equally likely to be reused. The purpose of this study was to shed light on how the speakers' conceptualisations of the referents under discussion affect reuse (along with a generation effect in memory documented in previous studies on dialogic reuse). Two experiments were conducted in which participants interactively added references to their common ground. From each participant's point of view, these references either did or did not match their own conceptualisation of the referents discussed, and were either self- or partner-generated. Although self-generated references were more readily accessible in memory than partner-generated ones (Experiment 1), reference reuse was mainly guided by conceptualisation (Experiment 2). These results are in line with the idea that several different cues (conceptual match, memory accessibility) constrain reference reuse in dialogue.

Keywords: dialogue; common ground; referential communication; memory; generation effect; conceptualisation

Introduction

Dialogue is a joint activity during which at least two partners collaborate in order to reach a common goal such as making dinner or working together (Bangerter & Clark, 2003; Clark, 1996). As they interact, speakers produce referring expressions, or references, to designate the objects and entities under discussion (“the living room table”, “the recipe I’m planning on making tonight”, “the boss”, “a timeslot which wouldn’t be very convenient for you”, etc.; e.g., Bard, Hill, Foster, & Arai, 2014; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fukumura, 2015). There are generally several different ways of talking about the same referent. For instance, the same table might be referred to as “the living room table” or “the beautiful wooden table I bought when I first moved into this place”. Speakers thus need to make a number of decisions about how to refer to things throughout the interaction.

In some cases, speakers refer to a referent several times during an interaction. When this happens, speakers reuse the same references, inferring that their partners should be capable of understanding them again (Brennan & Clark, 1996; Metzinger & Brennan, 2003). Such inferences require high-level perspective-taking skills. However, reference reuse also depends on lower-level processes such as whether these references are readily accessible in memory to them (e.g., Horton & Gerrig, 2005b; Knutsen & Le Bigot, 2014). Another factor which is likely to play an important role here is the way in which speakers personally view, or conceptualise, the referents under discussion. For instance, a piece of modern art may be viewed as “a weird painting” or as “an enlightened representation of humanity”. How do such different conceptualisations affect reference reuse? The current study seeks to address this question, thus contributing to a better understanding of referential decisions in dialogue.

Conceptual Pacts in Dialogue

According to the collaborative approach to dialogue, decisions about how to refer to things are adaptive: each speaker favours the production of references he or she believes to be easily understandable for his or her current addressee(s) (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fukumura, 2015; Isaacs & Clark, 1987; Nückles, Winter, Wittwer, Herbert, & Hübner, 2006; Yoon & Brown-Schmidt, 2014). For instance, Speaker A and Speaker B are planning an afternoon out. A wants to meet B in front of a shopping centre, but B does not know where this centre is. In this situation, A could adapt his or her speech by locating the shopping centre in relation to another landmark A knows B is capable of locating. For instance, if A and B have already gone to a nearby cinema together, A could refer to the shopping centre as “the centre which is very close to that cinema we went to”. To determine what their current partners are capable of understanding, speakers rely on their common ground, which includes the knowledge that dialogue partners share and are aware of sharing (Brown-Schmidt, 2012; Clark, 1996; Clark & Marshall, 1978, 1981; Stalnaker, 1978). The common ground includes information relative to the partners’ past shared experiences (e.g., the fact that A and B went to the cinema together in the example above) as well as shared community memberships (e.g., the fact that A and B both belong to a local cinema lovers community).

Past decisions about how to refer are also added to the common ground through a joint contribution process (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). This process is divided into two phases. During the first phase, one of the speakers presents a reference he or she believes to be suited to his or her current addressee’s understanding (e.g., A presents the reference: “the centre which is very close to that cinema we went to”). This reference is then accepted by the addressee, with the latter producing some kind of evidence that the reference presented was understood well enough for current purposes (e.g., B accepts this reference by saying: “oh yes, I know where that is”). A

reference may be accepted more or less explicitly: the addressee might accept it by nodding his or her head, by saying “okay”, “hm”, by moving on to the next relevant speech turn or by repeating the reference presented. It may also be accepted more or less immediately after the presentation phase, depending on whether the addressee has enough knowledge and information to understand it correctly straight away or whether he or she needs additional information to understand the director’s initial intentions. For instance, rather than accepting A’s reference immediately, B might reply: “I don’t know where that is, I really can’t remember where that cinema was”. In such situations, A might provide more information to B to help him or her to understand the reference initially presented (e.g., “it’s not far away from High Street”). Another possibility is that B might present an alternative reference (e.g., “do you mean the shopping centre that’s just in front of the town hall?”). In any event, at some point during this process, A and B will reach mutual understanding: they will converge on a reference which they are both capable of understanding and which B can finally accept.

Once presented and accepted, a reference is deemed part of the partners’ common ground. If either partner needs to refer to the same referent again later, he or she is likely to favour the reuse of the same reference in order to improve addressee comprehension. For instance, once the reference “the centre which is very close to that cinema we went to” (or any alternative reference finally accepted by B) has been added to A and B’s common ground, both A and B can assume that their partner should be capable of understanding this reference successfully if they reuse it again later during the interaction.

The contribution process was illustrated in a study conducted by Clark and Wilkes-Gibbs (1986) in which pairs of participants took part in a matching game (see also Fox Tree & Clark, 2013; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Hupet, Seron, & Chantraine, 1991; Krauss & Weinheimer, 1966; Lysander & Horton, 2012; Yoon & Brown-Schmidt, 2014). One of the participants (the director) described abstract pictures (tangram

figures; see Figure 1 for an example) to his or her partner (the matcher) in order to allow the latter to arrange these pictures in a predefined order. The same task was repeated six times. The main result was that the participants consistently used similar references to refer to the same pictures across trials (e.g., “the ice skater” to describe the picture presented in Figure 1), relying on the common ground built during the first trial(s) to infer that the matcher should be capable of understanding these references again in subsequent trials.

Building on these initial findings, Brennan and Clark (1996) suggested that dialogue partners rely on presentation and acceptance to establish conceptual pacts, that is, temporary agreements about how to refer to referents. For instance, in Clark and Wilkes-Gibbs's (1986) study, the director and the matcher agreed to refer to the picture presented in Figure 1 as “the ice skater”; the director tended to reuse this reference (rather than another, potentially suitable different reference) across trials. Such pacts are defined as temporary because when speakers face new addressees, they must establish new (and potentially different) conceptual pacts. For instance, if the director started a new interaction with a different matcher, he or she might use another reference than “the ice skater” to refer to the picture in Figure 1. Conceptual pacts also affect comprehension, as speakers breaking such pacts (i.e., using a new reference rather than reusing a reference that belongs to the common ground) impairs comprehension for addressees (Metzing & Brennan, 2003; see also Yoon & Brown-Schmidt, 2014).

Insert Figure 1 around here

Biases in Reference Reuse

The findings reviewed in the previous section suggest that speakers tend to reuse the same references when referring repeatedly to the same referent in order to adapt to their partners. However, there is evidence that not all presented and accepted references are equally likely to be subsequently reused in the remainder of the interaction. To illustrate this point, Knutsen

and Le Bigot (2014) ran an experiment in which pairs of participants added references to landmarks featured on a map to their common ground through presentation and acceptance. The main finding was that speakers were more likely to reuse self-presented references than partner-presented ones (i.e., to re-refer to landmarks corresponding to self-presented references more than to landmarks corresponding to partner-presented references). A similar finding was obtained in another study in which the participants had the opportunity to generate references to landmarks rather than simply read them off a map (Knutsen, Ros, & Le Bigot, in press). Based on this finding, the authors have suggested that reference reuse is subject to a *self-presentation bias*. However, these studies did not focus on decisions about how to refer per se. Indeed, when a speaker reuses a reference, he or she must make two different decisions, the first of which pertaining to whether to mention a referent or not, and the second of which pertaining to how to refer to this referent. To illustrate this distinction, consider the following example. A is discussing a route with B. At some point, A presents the following building to refer to the town hall: “this horrible building in the middle of the town”. B accepts this reference, thus agreeing to establish this conceptual pact and to add this reference to the common ground. Later during the interaction, A has the opportunity to refer to the same building again. At this point, A first needs to decide whether or not to mention this referent again. Such decisions are hereafter referred to as *decisions to refer* (i.e., decisions about *whether* to refer or not to a given referent, regardless of how this referent is actually referred to). Actual *decisions about how to refer* only take place after this, once A has decided to mention the referent. At this point, A must decide whether to maintain the conceptual pact by reusing the reference which was initially agreed upon or to abandon it by using a different reference.

Knutsen and colleagues (Knutsen & Le Bigot, 2014; Knutsen et al., in press) only examined decisions to refer, showing that the self-presentation bias makes speakers more

likely to decide to mention a referent again when the reference to this referent was initially self-presented than when it was initially partner-presented. However, a similar self-presentation bias might also constrain actual decisions about how to refer, making speakers more likely to resort to self-initiated conceptual pacts than to partner-initiated conceptual pacts. For instance, if Speaker A refers to a table as “an old piece of furniture” and Speaker B refers to the same table as “a beautiful wooden table”, the self-presentation bias could make A more likely to reuse the reference “an old piece of furniture” and B more likely to reuse the reference “a beautiful wooden table” when referring to this table again. The first purpose of this study is to address this possibility, thereby offering a better understanding of decisions about how to refer in collaborative dialogue.

The second goal of the current study is to offer a better understanding of the processes underlying a potential self-presentation bias in decisions about how to refer. Knutsen and Le Bigot (2014) have suggested that the self-presentation bias in decisions to refer is due to a self-production effect in memory, whereby words produced aloud are remembered better than words read silently, especially when these words are self-produced (as opposed to partner-produced; MacLeod, 2011; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). In cases where speakers generate the references at the time of presentation (Knutsen et al., in press), this effect could also be due to a generation effect in memory (e.g., Burnett & Bodner, 2014; Rosner, Elman, & Shimamura, 2013; Slamecka & Graf, 1978). These suggestions are corroborated by the finding that after the end of an interaction, speakers tend to remember self-generated content better than partner-generated content (Hjelmquist, 1984; Jarvella & Collas, 1974; Knutsen et al., in press). They are also corroborated by the finding that other occurrences of reference production during the interaction (e.g., acceptance through verbatim repetition as well as reuse per se) increase subsequent accessibility in memory (Knutsen & Bigot, 2014, 2015). This is in line with the idea that “ordinary” (i.e., non-specific to dialogue)

low-level processes play an important role in reference production in dialogue (Gorman et al., 2013; Horton, 2007, 2008; Horton & Gerrig, 2002, 2005a, 2005b).

Regarding decisions about how to refer, the potential tendency to reuse self-presented references more often could be due not only to memory, but also to additional determinants. In particular, it could be due to the way in which each speaker personally conceptualises, or “views”, the referents under discussion. For instance, if a speaker refers to the Tangram figure presented in Figure 1 by presenting the reference “the ice skater”, this usually means that the speakers “sees” an ice skater when he or she looks at this figure. Importantly, dialogue partners might not necessarily conceptualise the same referents in the same way. For instance, when Speaker A refers to the table mentioned previously as “an old piece of furniture” and Speaker B refers to the same table as “a beautiful wooden table”, the two references produced reflect two different conceptualisations of the same referent. (It is noteworthy that the notion of “conceptualisation” in the current study can be linked with the perspective-taking literature, and more particularly with the notion of Level-2 perspective-taking, which also refers to how people “view” a scene (e.g., Flavell, Everett, Croft, & Flavell, 1981). However, in order to link the current work to theoretical models of language production (see the General Discussion), the notion of “conceptualisation” is preferred here.)

Self-presented references being reused subsequently might therefore simply reflect the fact that speakers tend to “view” the same referents in the same way in addition to, or rather than, better memory for self-presented references¹. One implication would be that in

¹ This suggestion should nonetheless be qualified in light of the fact that adding references to the common ground through presentation and acceptance is a joint process in which both partners play an active role: the partner performing the acceptance phase may suggest revising the reference initially presented (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). This point will be addressed in the General Discussion.

cases where Speaker A presents a reference which happens to match Speaker B's conceptualisation of the referent under discussion, Speaker B could become more likely to reuse a partner-presented reference - a prediction incompatible with the (self-) production/generation effect account (Knutsen & Le Bigot, 2014; Knutsen et al., in press). In other words, there would be no such thing as a single "self-presentation effect" in such cases. Rather, generation and conceptualisation would differentially affect reference reuse.

The idea that conceptualisation is an important determinant of reference reuse was addressed by Duff, Hengst, Tranel, and Cohen (2006; note that the influence of self- vs. partner-presentation was not assessed in this study). These authors compared decisions about how to refer in participants with amnesia (hippocampal amnesia; see Spiers, Maguire, & Burgess, 2001) and typical participants. The experimental setting was similar to that used by Clark and Wilkes-Gibbs (1986): the participants played a matching game which required referring to abstract Tangram figures. The analysis of the results revealed that just like typical participants, participants with amnesia reused the same references to refer repeatedly to the same referents when they played the role of directors. Such reuse could not be due to the fact that these participants remembered which references had initially been presented, as hippocampal amnesia prevents individuals from encoding new information in episodic memory. This was corroborated by the results of a follow-up test in which tangram figures (which had not been used in Duff et al.'s main experiment) were arbitrarily paired with nouns (e.g., a tangram figure looking like a rabbit was paired with the word "broom"). The participants were asked to learn these pairs. The analysis of their performance on a subsequent memory test revealed that the performance of the participants with amnesia was strongly impaired. On the basis of these results, the authors suggested that reference reuse during the matching task did not reflect the amnesic participants' memory for references previously presented and accepted, but rather that the participants conceptualised the pictures

in the same way across trials, leading them to produce similar descriptions every time they encountered the same referent.

In sum, reference reuse in dialogue could be guided by a production/generation effect (Knutsen & Le Bigot, 2014), a “conceptual match” effect (Duff et al., 2006), or both. In many cases, these two effects are likely to be confounded, as it seems reasonable to assume that speakers often generate references which match their own conceptualisations at the time of initial reference presentation. This makes it particularly difficult to examine the respective contributions of conceptualisation and generation to dialogic reference reuse. The current study sought to address this question in a situation where participants made no decisions to refer (i.e., they did not decide which referents to refer to during the task; this was decided in advance by the experimenter), but had the opportunity to make decisions about *how* to refer as they interacted repeatedly about the same referents.

Current Study: Overview and General Hypothesis

The general hypothesis tested in this study was that reference reuse in dialogue is guided not only by a generation effect in memory (i.e., speakers reuse self-generated references more often than partner-generated ones), but also by a conceptual match effect (i.e., speakers reuse references which match their conceptualisations more than references which do not match their conceptualisations). In two experiments, pairs of participants added references to abstract Tangram figures to their common ground through presentation and acceptance. In some cases, the participants generated references which matched their own conceptualisations; in other cases, the participants were forced to generate references which did not match their own conceptualisations.

To test the general hypothesis, the study was divided into two steps. The purpose of the first step (Experiment 1) was to confirm that the generation effect affects not only

speakers' memory for which referents were *mentioned* during an interaction (as shown in previous studies on the self-presentation bias), but also their memory for which references were actually *used* to refer to these referents – or, in other words, their memory for conceptual pacts (Brennan & Clark, 1996; Metzinger & Brennan, 2003). The purpose of the second step (Experiment 2) was to examine how the generation effect and the conceptualisation match effect found in Experiment 1 affect reference reuse in dialogue.

Experiment 1

Rationale and Operational Hypotheses

Experiment 1 was divided into three phases. During the first phase (Dialogue Phase), pairs of participants were shown various tangram figures one by one. For each figure, each participant was asked to generate a reference and to answer any questions his or her partner might have about his or her reference. Thus, each pair generated two references for each figure discussed.

For each figure, the participants had to generate a different reference; each participant went first on half the trials and second on half the trials. Therefore, in some cases, each participant had the opportunity to present a reference that matched his or her conceptualisation of the picture (i.e., when he or she went first or when he or she went second but that the reference presented by his or her partner did not match the participant's own conceptualisation); in other cases, the participant was forced to present a reference that did not match his or her own conceptualisation (i.e., when he or she went second and that the reference matching his or her conceptualisation had already been presented by his or her partner). In other words, each participant generated references which matched or did not match his or her conceptualisations and heard his or her partner generating references which matched or did not match his or her (i.e., the participant's) conceptualisations. During the second phase of the experiment (Memory Assessment Phase), the participants were asked to

recall (in writing) the two references generated for each tangram figure during the Dialogue Phase. This allowed for examination of whether the participants' memory for the references generated during the Dialogue Phase was subject to a self-generation bias. Finally, during the third phase of the experiment (Conceptualisation Assessment Phase), the participants were asked to write down, for each figure used during the Dialogue Phase, which of the two references generated matched their conceptualisation best. Doing so allowed for determination of whether the self-generation bias found in the Memory Assessment Phase held for references which matched the participants' conceptualisations and also for references which did not.

The main operational hypothesis was that self-generated references are more likely to be recalled than partner-generated ones. The second operational hypothesis was that the self-generation effect affects both references which match the partners' conceptualisation and references which do not match the partners' conceptualisation.

Participants

Forty native French speaking students (22 female; mean age 18.75, $SD = 1.28$) took part in the experiment for course credit. The participants signed an informed consent form before the beginning of the experiment and were fully debriefed after the experiment.

Apparatus

The dialogues between the participants were recorded using a TASCAM DR-40 double-entry digital recorder connected to two lapel microphones (one per participant).

Materials

A hundred randomly selected tangram figures were used in this experiment (Figure 2). These figures came from various tangram sets. Each tangram figure was randomly assigned a number between 1 and 100; the figures were then randomly divided into five groups of 20 figures labelled A, B, C, D and E (i.e., Group A was made of figures 1 through 20, Group B was made of figures 21 through 40, etc.; the 20 figures discussed by a pair of participants during the Dialogue Phase all belonged to the same group). Each figure was printed on a separate A6 sheet of paper for use during the Dialogue Phase.

Insert Figure 2 around here

Paper booklets were then prepared for each participant to use during the Memory Assessment Phase. Each booklet included 20 target tangram figures (i.e., the 20 figures discussed during the Dialogue Phase) as well as 20 distractor tangram figures which belonged to another group of figures (e.g., when the target figures belonged to group A, the distractors belonged to group B). Two different versions of these booklets were created (the random order in which the figures were presented in the booklet was different in each booklet). Two lines were printed in front of each figure so that the participant could write down the two references generated during the Dialogue Phase.

Finally, the same pictures as in the Dialogue Phase as well as blank lined A5 sheet of paper were used in the Conceptualisation Assessment Phase.

Task and Procedure

The experiment was performed by pairs of participants who sat next to each other and facing the experimenter in a quiet experimental room. Before the beginning of the experiment, the participants were informed that the study sought to investigate referential communication and that the experimenter was simply interested in how speakers refer to abstract pictures such as tangram figures. They were also told that the experiment would involve more than one phase

but were not informed in advance of how many phases there would be or of the content of the different phases.

The Dialogue Phase was divided into 20 trials (see Table 1 for an example of a dialogue between two participants during a trial; in this example, the participants were describing the figure presented in Figure 3). At the beginning of each trial, the experimenter showed the participants a tangram figure (both participants were shown the same picture; because the participants sat next to each other, they both viewed the figure from the same perspective). Two random presentation orders were used in this phase. One of the participants (hereafter referred to as Participant X) generated a reference to designate the figure. The participants were then free to discuss Participant X's decision: for instance, Participant Y could ask for additional information or manifest his or her understanding by saying "okay". For instance, in Table 1, Participant X started by presenting the reference "someone on a boat". This presentation was followed by a brief discussion during which X provided Y with additional information ("you have the boat here and there you have the guy") and during which Y manifested his or her understanding ("yes", "I see"). Y could also have used this discussion to ask X for additional information (e.g., "I don't understand, if it's a boat, where is the hull?") or express a judgement about X's reference (e.g., "I don't agree at all"). Once Participants X and Y had finished discussing Participant X's decision, Participant Y proposed a different reference to the same figure and then briefly discussed his or her decision with Participant X. The first reference was presented by Participant X in 50% of the trials and by Participant Y in the other 50% of trials. At the beginning of each trial, the experimenter prompted the participant whose turn it was to present the first reference to start talking. This participant had as much time as necessary before he or she actually started talking.

Insert Figure 3 around here

Insert Table 1 around here

Then, during the Memory Assessment Phase, each participant performed an individual memory test. To this end, each participant was given a booklet and a pen. For each tangram figure shown in the booklet, the participant was asked to do two things: (1) to decide whether or not it had been shown during the Dialogue Phase and (2) if this was the case, to attempt to recall all of the words presented by each participant to describe this figure. The experimenter stayed in the room with the participants to make sure that they did not attempt to communicate with each other during this individual phase.

Finally, the participants' conceptualisations were assessed during the Conceptualisation Assessment Phase. During this phase, the participants were shown again the 20 tangram figures they had been shown during the Dialogue Phase. These were presented one by one in the same order as in the Dialogue Phase; this was to make sure that the amount of time elapsed between the presentation of a figure during the Dialogue Phase and the presentation of the same figure during the Conceptualisation Assessment Phase was roughly the same for all figures. For each figure, each participant was first asked to say aloud the reference that he or she had initially presented during the Dialogue Phase. If one of the participants could not remember the reference which he or she had presented during the Dialogue Phase, his or her partner was allowed to help him or her by producing the reference him- or herself if he or she could remember it (this happened in less than 2% of trials, suggesting that the participants remembered well the references they had generated during the Dialogue Phase). The participants were allowed to interact at this point. The purpose of this was to make sure that both references were readily accessible in memory to both participants. Each participant's conceptualisation was then assessed by asking him or her to write down which of these two references reflected his or her point of view better (the participants were instructed not to interact as they wrote down their answer and not to tell their partner which reference they had selected). The participants were specifically required to choose between

the two references initially generated during the Dialogue Phase: they could not use a new reference during this phase.

The participants performed the three phases of the experiment at their own pace. There was no break between the three phases. The experiment lasted approximately one hour. The participants were fully debriefed after the end of the Conceptualisation Assessment Phase. The three steps of the experiment are summarised in Table 2.

Insert Table 2 around here

Data Coding and Experimental Design

Coding – generation. The interactions between the participants during the Dialogue Phase were transcribed and the content words (proper and common nouns, adjectives and verbs) presented to refer to the tangram figures were identified. Content words included proper nouns (e.g., “San Francisco”), common nouns (e.g., “cat”), adjectives (e.g., “tall”) and verbs (e.g., “to eat”). Auxiliary verbs (i.e., “to be” and “to have”), modal verbs (e.g., “can”, “must”), determiners (e.g., “the”, “a”, “one”), pronouns (“I”, “this”), adverbs (e.g., “often”), prepositions (e.g., “after”, “despite”), coordination conjunctions (e.g., “but”, “and”), disfluencies (e.g., “uh”, “hm”) and interjections (e.g., “pew”, “oh”) are not content words and were therefore not identified in the corpus. In addition, only the content words used to describe the tangram figures were taken into account here: the content words referring to the participants’ perception of the figures (e.g., “I see...”) were not taken into account. For instance, in the example in Table 1, the content words “boat”, “guy” and “hat” were coded as presented, but the words “well” (interjection), “uh” (disfluency) or “the” (determiner) were not coded, as they are not content words. In a similar way, the verb “to see” was not coded as presented, as it referred to Participant X’s perception of the figure. Only the first mention of each content word (for each figure presented) was taken into account.

In some cases, one of the participants would present a reference and the other participant would complete this description. For instance, Participant X would say “I see a man on a boat” and Participant Y would say “yes, he seems very happy”. In such cases, the extra content words presented by the other participant (in this example, “happy”) were not taken into account. Indeed, the participants’ task was to generate two different references for the figures they were shown (and not to complete each other’s descriptions); therefore, the coding only sought to identify which content words were produced by Participant X in his or her descriptions on one hand and which content words were produced by Participant Y in his or her descriptions on the other. This led us to remove 424 content words from the dataset.

Each content word was then coded depending on whether it had been self-generated or partner-generated from each participant’s point of view. For instance, in the example shown in Table 1, the content words “boat” and “guy” were coded as self-generated from X’s point of view and partner-generated from Y’s point of view; as for the content word “hat”, it was coded as self-generated from Y’s point of view and partner-presented from X’s point of view. This level of coding was used as one of the predictors in the main statistical analysis.

Coding – conceptualisation. The participants’ responses during the Conceptualisation Assessment Phase were examined to determine, for each tangram figure, which reference presented matched each participant’s conceptualisation. This allowed for determination of whether the references generated by each participant did or did not match his or her conceptualisation. For instance, if both Participant X and Participant Y indicated that they viewed the tangram figure described in Table 1 as “a boat” during the Conceptualisation Assessment Phase, this word was coded as matching one’s conceptualisation best from both participants’ point of view. All of the other content words presented to describe this picture (i.e., “guy”, which was initially presented by Participant X,

and “hat”, which was initially presented by Participant Y) were coded as not matching one’s conceptualisation from X’s point of view and from Y’s point of view. This level of coding was used as one of the predictors in the main statistical analysis.

Coding – recall. The participants’ memory for the references presented during the Dialogue Phase was assessed by examining their performance during the Memory Assessment Phase. For each participant and each tangram figure, each content word presented during the dialogue phase was coded either as recalled (code 1) or non-recalled (code 0). This was a binary variable; however, this variable also reflected the proportion of content words recalled. For instance, if Participant X presented two content words and Participant Y presented three content words during the Dialogue Phase and that Participant X recalled four of these content words, this resulted in four out of five content words being coded 1 for recall, which corresponded to an average recall proportion of 0.8.

This binary level of coding was used as the dependent variable in the main statistical analysis.

Independent variables. There were two independent variables in this experiment. The first one was Generation. From each participant’s point of view, each content word presented was either self-generated or partner-generated. The second one was Conceptualisation. From each participant’s point of view, each content word presented either match or did not match his or her own conceptualisation. Both variables were within-participants.

Results

Descriptive statistics – Dialogue Phase. During the Dialogue Phase, the average number of speech turns produced per dyad was 132.70 ($SD = 50.81$) and the average number of words produced per dyad (regardless of whether these were content or non-content words) was 1307.65 ($SD = 555.23$).

The total number of content words presented by the participants to describe the tangram figures was 3241 (note that this figure does not include the content words presented by one participant to complement a description initiated by the other participant). The average number of content words produced per tangram figure per participant was 4.05 ($SD = 2.63$).

One of the participants asked his or her partner for more information after a reference had been generated in 18.50% of trials (74 trials out of 20 pairs x 20 figures = 400 trials). Thus, clarification requests were not systematic; nonetheless, this confirms that the participants viewed this phase as interactive and felt that they could ask for additional information when necessary.

Descriptive statistics – Memory Assessment Phase. During the Memory Assessment Phase, fillers were incorrectly identified as having been discussed during the Dialogue Phase in 1.38% of cases (11 incorrect identifications out of 20 fillers x 40 participants = 800 occurrences). Target figures were correctly identified as having been discussed during the Dialogue Phase in 91.88% of cases (735 correct identifications out of 20 target figures x 40 participants = 800 occurrences). This suggests that the participants remembered well which referents had been discussed during the Dialogue Phase.

A total of 1800 content words which had been discussed during the Dialogue Phase were recalled during the Memory Assessment Phase (the content words corresponding to tangram figures which had not been discussed during the Dialogue Phase and the content

words which had not been produced during the Dialogue Phase, although they were recalled in association with a figure which had been discussed during this phase, were discarded from further analysis, as the hypotheses solely concerned previously produced references to known tangram figures). The average number of content words correctly recalled per tangram figure per participant was 2.25 ($SD = 1.30$).

Descriptive statistics – Conceptualisation Assessment Phase. During this phase, pairs of participants were able to recall the two references initially presented in 98.63% of cases (789 occurrences out of 20 pictures x 40 participants = 800 occurrences). The trials corresponding to the remaining 1.37% (11 trials) were discarded from further analysis, as these corresponded to situations in which the two references were not necessarily readily available for the two participants to choose from. In a similar way, trials in which the participants' responses included both self- and partner-generated words (8.88% of cases: 71/800 occurrences) were also excluded from further analysis, as these reflected cases in which the participants had no clear preference as to how to conceptualise the referents. The remaining dataset included data from 718 trials, representing a total of 2895 presented words.

The participants reported that the first reference presented matched their conceptualisation best in 60.86% of cases (437/718 occurrences). This confirms that the first reference presented during the Dialogue Phase sometimes matched the participants' conceptualisation even when this reference was not self-generated, as the participants only generated the first reference in 50% of trials. They also reported that the reference that they had generated themselves matched their conceptualisation best in 60.31% of cases (433/718 occurrences). This pattern of results is summarised in Table 3.

Insert Table 3 around here

Main analysis: Influence of generation and conceptual match on reference recall during the Memory Assessment Phase. A logistic mixed model was used to analyse the data (for logistic analysis, see Agresti, 2002; Jaeger, 2008; for mixed models, see Barr, Levy, Scheepers, & Tily, 2013).

Mixed models allow for the inclusion of random intercepts (which are used to account for potential variability across dyads, across participants, and across items) and for the inclusion of random slopes (which are used to account for the fact that different dyads, different participants and different items might differ in their sensitivity to the fixed effects included in the model). Mixed models are also used to account for the nesting of participants in larger groups such as dyads (see McMahon, Pouget, & Tortu, 2006). The current analysis was conducted using SAS 9.4 (GLIMMIX procedure). In line with Barr et al.'s (2013) recommendations, the maximal random structure justified by the experimental design (i.e., all random intercepts and all random slopes corresponding to within-unit IVs) was initially implemented. Three levels of random effects were distinguished in this analysis: dyads, participants and tangram figures (items). Then, in line with Kiernan, Tao, and Gibbs (2012), the random effects causing G-matrix convergence failure were identified and removed from the model. The identification of problematic random effects is performed automatically in SAS. Removing these random effects from the model has no effect on the outcome of the analysis (i.e., the model parameters remain unchanged; Kiernan et al., 2012; Knutsen & Le Bigot, 2014). The results reported hereafter correspond to the final model (i.e., the model in which the random effects causing convergence problems were not included).

As for *logistic* mixed models, they are used in situations where the outcome of the analysis is binary (e.g., in this experiment, any content word was either coded as recalled or non-recalled). One of the parameters returned by logistic models is the odd ratio (OR), which compares the odds associated with two different events (Jaccard, 2001). For instance, the

current study compared the odds of recalling self-generated content words and the odds of recalling partner-generated content words during the Memory Assessment Phase. If the latter were used as the baseline in the analysis and that an OR of 1.5 was found, this would mean that 1.5 self-generated content words were recalled for every one partner-generated content word. The corresponding 95% confidence interval is also usually reported. ORs are also informative with regard to effect size: larger odd ratios reflect larger effects (Agresti, 2002).

The number of content words presented in each cell of the design varied across cells, making it difficult to assess the degrees of freedom should be used to determine whether or not the effects involved in this analysis were statistically significant. In such cases, the Satterthwaite approximation may be used to correct the degrees of freedom, which is what was done in the current study (Keselman, Algina, Kowalchuk, & Wolfinger, 1999; Satterthwaite, 1946).

The model used to analyse the data included Generation (self, other), Conceptualisation (match, mismatch) and the interaction between these two factors as fixed effects. The outcome variable was the likelihood of recalling a content word during the Memory Assessment Phase. The random effect structure included (1) by-dyad random intercepts and by-dyad random slopes corresponding to Conceptualisation; (2) by-participant random intercepts and by-participant random slopes corresponding to Generation; (3) by-item random intercepts and by-item random slopes corresponding to Conceptualisation. All other random effects (i.e., by-dyad random slopes corresponding to Generation, by-participant random slopes corresponding to Conceptualisation and by-item random slopes corresponding to Generation) were removed from the analysis because they caused model convergence failure. Removing these random effects had no influence on the outcome of the analysis.

The results are reported in Figure 4. Generation significantly predicted recall, $F(1, 67) = 9.59, p = .003$. The participants were more likely to recall self-generated content words

than partner-generated ones, $OR = 1.43$, $CI_{.95} = 1.14, 1.80$. Conceptualisation also significantly predicted recall, $F(1, 22) = 350.09$, $p < .001$. The participants were more likely to recall content words which matched their conceptualisation than content words that did not match their conceptualisation, $OR = 16.34$, $CI_{.95} = 11.99, 22.28$. Furthermore, the interaction between these two factors was statistically significant, $F(1, 5793) = 19.67$, $p < .001$. Additional pairwise comparisons (Sequential Bonferroni) were conducted to offer a better understanding of this pattern of results. These comparisons revealed that the difference between self- and partner-generated references was significant when these references matched the participants' conceptualisation ($p < .001$) but not when these references did not match the participants' conceptualisations ($p = 1.00$). The model parameters are reported in Appendix A (Table A1).

Insert Figure 4 around here

As mentioned previously, generation at the time of presentation is not the sole known linguistic determinant of reference accessibility in memory in dialogue. Reference reuse also depends on how these references were initially accepted: references accepted through verbatim repetition are more likely to be reused in the remainder of the interaction (Knutsen & Le Bigot, 2012, 2014, 2015; Knutsen et al., in press). Furthermore, references reused during an interaction are remembered better after the end of the interaction than references which were not reused (Knutsen & Le Bigot, 2014, 2015; Knutsen et al., in press). In the current study, this kind of reused is labelled "short-term reuse" in order to distinguish it from longer-term reuse (i.e., reuse occurring in subsequent interactions between the same partners, such as reuse during the Matching Phase in Experiment 2).

These two determinants of reference reuse were not of prime interest in the current study; furthermore, the number of acceptances and short-term reuses in this study was too small to include these as fixed effects in the model. Nonetheless, acceptance and short-term

reuse could have affected the participants' memory performance during the Memory Assessment Phase. To discard this eventuality, the participants' interactions during the Dialogue Phase were coded for acceptance and short-term reuse (i.e., reuse during the Dialogue Phase); a statistical analysis was then performed to determine whether the effects of Generation and Conceptualisation found in the main analysis remained significant when Acceptance and Reuse were controlled for. This analysis is reported in Appendix B: it confirmed that the effects of Generation and Conceptualisation on recall remained significant even when these additional determinants were taken into account.

Discussion

The purpose of Experiment 1 was to determine whether speakers' memory for decisions about how to refer is subject to a generation effect (Slamecka & Graf, 1978). The results corroborated this idea, as the participants recalled the content words they had generated themselves better than the content words which had been generated by their partner. This confirms and extends previous work on how production and generation affect collaborative dialogue (Knutsen & Le Bigot, 2014; Knutsen et al., in press), shedding light on how low-level processes affect speakers' memory for past interactions (see also Knutsen & Le Bigot, 2015).

Conceptual match was also taken into account in Experiment 1. The results revealed that content words were more likely to be recalled when they matched the participants' conceptualisations than when they did not. However, this effect did not necessarily reflect better memory for these words. Indeed, one possibility is that the participants perceived the figures they were shown during this phase in the same way as in the Dialogue Phase, thus causing them to use the same words to describe them again in an individual task (see Duff et al., 2006).

Moreover, the interaction between Generation and Conceptualisation revealed that the self-generation effect was only statistically significant when the references recalled matched one's own conceptualisations. This does not necessarily imply that references which do not match one's own conceptualisations are not subject to a generation effect as well. Indeed, no conclusions can be drawn from the lack of a significant difference between self- and partner-generated references in this condition. However, this suggests that the generation effect is attenuated for references which do not match one's own conceptualisations. This could be due to a floor effect in this experiment: the participants might have had difficulty recalling the references which did not match their own conceptualisations, regardless of who had generated them. In any event, the significant interaction also reflects the fact that the participants' tendency to recall better references which matched their own conceptualisations was stronger when these references had initially been self-generated (rather than partner-generated). Thus, references which benefit from both a generation effect and a conceptual match effect are remembered better than any other references.

In sum, Experiment 1 suggests that past decisions about how to refer are subject to a generation effect – at least in cases where these references match one's conceptualisations of the referents under discussion. The purpose of Experiment 2 was to build on this initial finding by examining how these two factors affect actual reference reuse in dialogue. To this end, in Experiment 2, the initial Dialogue Phase was followed by a matching game instead of a memory test. This game gave the participants the opportunity to reuse the references presented during the Dialogue Phase, allowing us to examine how generation and conceptualisation affect reference reuse.

Experiment 2

Rationale and Operational Hypotheses

Just like Experiment 1, Experiment 2 was divided into three phases. The first phase was identical to the Dialogue Phase in Experiment 1. Then, the participants embarked on a Matching Phase during which they took turns describing tangram figures to each other (in target trials, these figures had already been discussed during the Dialogue Phase). Finally, the participants' conceptualisations were assessed in a third phase, which was identical to the Conceptualisation Assessment Phase in Experiment 1.

Two operational hypotheses about reference reuse during the Matching Phase were formulated. The first hypothesis was that if reuse during the Matching Phase is mainly guided by generation (as suggested by Knutsen & Le Bigot, 2014), the participants will reuse self-generated references more often than partner-generated ones regardless of whether these references match their conceptualisation or not. By contrast, the second hypothesis was that if reuse during the Matching Phase depends both on generation and conceptual match, the participants will reuse self-generated references more often than partner-generated references *and* references that match their conceptualisations more often than references that do not.

Participants

Forty participants (35 female; mean age 18.40, $SD = 0.93$) were recruited under the same conditions as in Experiment 1 to take part in Experiment 2.

Apparatus

The same digital recorders as in Experiment 1 were used to record the interactions between the two participants in the first two phases of the experiment.

Materials

The same tangram figures as in Experiment 1 were used in Experiment 2. As in Experiment 1, each figure was printed on a separate A6 sheet of paper for use during the Dialogue Phase.

Paper booklets were prepared for use during the Matching Phase (see Figure 5). Each booklet was 40 pages long. Four different tangram figures were printed on each page. One of these figures was one of the figures discussed during the Dialogue Phase (each of the figures discussed during the Dialogue Phase thus appeared twice in each booklet) and the other three belonged to a different group (e.g., the booklets prepared for the dyads who initially discussed the figures from group A also included the figures from groups B, C and D). The position of the figure discussed during the Dialogue Phase among the other three figures was randomized. Two random page presentation orders were used to create these booklets. Five booklets were created depending on which group of figures (A, B, C, D or E) had been discussed during the Dialogue Phase.

Two complementary versions of each of these five booklets were then created (see Figure 5). The two versions of the same booklet included the same tangram figures printed in the same order. In the first version, on even-numbered pages (20 in total), the target figure was encircled by a red square; on odd-numbered pages, the target figure was not encircled (20 in total). In the second version of the same booklet, target figures were encircled on odd-numbered pages. In other words, when the target figure was encircled in one of the booklets, it was not encircled in the other booklet. In 50% of cases, the target figure was the figure discussed during the Dialogue Phase; in the other 50% of cases, the target figure belonged to another group. Pairs of complementary booklets were elaborated so that each figure discussed during the Dialogue Phase served as target once and as distractor once. The purpose of this was to prevent the participants from inferring that when the array of figures included one of the figures discussed during the Dialogue Phase, the figure was necessarily the target during the Matching Phase.

Insert Figure 5 around here

The materials used during the third phase were identical to those used during the Conceptualisation Assessment Phase in Experiment 1.

Task and Procedure

As in Experiment 1, the participants were informed that the experiment would be divided into several phases, but they were not informed in advance of what they would be asked to do during each of these phases. The Dialogue Phase in this experiment was identical to the Dialogue Phase in Experiment 1 (see Table 4 for a dialogue sample in this phase; the figure discussed in this sample can be found in Figure 6).

Insert Figure 6 around here

Insert Table 4 around here

The Matching Phase was divided into 40 trials, each corresponding to a different page of the participants' booklets. In each trial, the task of the participant playing the role of director was to describe the tangram figure encircled by a red square to the participant playing the role of matcher so that the latter could find this figure among the four represented on his or her booklet and give the director the corresponding number (1, 2, 3 or 4, starting from the left). The trial ended after the director had confirmed that the matcher's answer was correct. The participants switched roles (Director and Matcher) after each trial, implying that each participant had the opportunity to describe 10 target figures and 10 filler figures to his or her partner during this phase (see Table 5 for two dialogue samples from this phase).

Insert Table 5 around here

Finally, the Conceptualisation Assessment Phase in this experiment was identical to the Conceptualisation Assessment Phase in Experiment 1. The participants performed the three phases of the experiment at their own pace and were fully debriefed after the end of the

Conceptualisation Assessment Phase. The experiment lasted approximately one hour. The three steps of the experiment are summarised in Table 6.

Insert Table 6 around here

Data Coding and Experimental Design

Coding – generation and conceptualisation. The data from the Dialogue Phase and the Conceptualisation Assessment Phase were coded for generation and conceptualisation in the same way as in Experiment 1. 532 words were removed from the dataset due to having been presented by one of the participants to complement the other participant’s description during the Dialogue Phase.

Coding – reuse during the Matching Phase. The content words produced by the Director during the Matching Phase were identified as in Experiment 1. For instance, in the examples given in Table 5, Director X produced the content words “boar” and “warthog” (Upper Panel) and Director Y produced the content words “UFO”, “diamond” and “square” (Lower Panel). The list of content words was then used to code each word generated during the Dialogue Phase as reused (code 1) or non-reused (code 0). For instance, in the example provided in Tables 4 and 5, the content word “warthog”, which was generated by Participant X during the Dialogue Phase, would have been coded as reused by this participant during the Matching Phase. All other content words initially generated during the Dialogue Phase (i.e., “boat” and “anchor”) would have been coded as non-reused. Just like recall in Experiment 1, reuse was a binary variable. In addition to this, the content words produced in fillers trials were not coded for reuse, as the figures discussed during these trials had not been discussed during the Dialogue Phase.

This level of coding was used as the dependent variable in the main statistical analysis.

Independent variables. Just like Experiment 1, Experiment 2 involved two within-participants independent variables, Generation and Conceptualisation. These were coded in the same way as in Experiment 1.

Results

Descriptive statistics – Dialogue Phase. During the Dialogue Phase, the average number of speech turns per dyad was 158.60 ($SD = 69.11$) and the average number of words produced by dyad (regardless of whether these were content or non-content words) was 1434.95 ($SD = 649.84$). The total number of content words presented by the participants was 3406 (note that this figure does not include the content words presented by one participant to complement a description initiated by the other participant). The average number of content words produced per tangram figure per participant was 4.26 ($SD = 2.85$).

One of the participants asked his or her partner for more information after a reference had been generated in 14.50% of trials (58 trials out of 20 pairs x 20 figures = 400 trials).

Descriptive statistics – Matching Phase. During the Matching Phase, the average number of speech turns per dyad was 156.10 ($SD = 39.90$) and the average number of words (including content and non-content words) produced per dyad was 861.60 ($SD = 373.12$). A total of 562 (non-necessarily unique, as part of the words presented within a dyad might have also been presented within one or several other dyads) content words which had been presented during the Dialogue Phase were reused by the Directors on target trials (i.e., trials in which the figure to describe had previously been discussed during the Dialogue phase; the

data corresponding to non-target trials were not analysed further, as no hypothesis was formulated concerning these trials). The average number of reused content words per tangram figure per director was 1.41 ($SD = 0.98$).

Matchers managed to find the target figure in their first attempt in 96.25% of trials (770 trials out of 20 dyads x 40 trials = 800 trials in total), suggesting that the task was relatively easy for the participants.

Descriptive statistics – Conceptualisation Assessment Phase. The trials in which the participants were not able to recall the two references initially presented and the trials in which the participants' responses included both self- and partner-generated words (15.38% of cases: 123 out of 800 occurrences) were excluded from further analysis. The remaining dataset thus included data from 677 trials, which represented a total of 2945 presented words.

The participants reported that the first reference presented matched their conceptualisation best in 59.68% of cases (404/677 occurrences). They also reported that the reference that they had generated themselves matched their conceptualisation best in 63.52% of cases (430/677 occurrences). This pattern of results is summarized in Table 7.

Insert Table 7 around here

This pattern of results is informative with regard to a potential confound in this study. Indeed, one might suggest that giving the participants the opportunity to reuse the references presented in the Dialogue Phase during the Matching Phase (and to assess the efficiency of such reuse on their partners' comprehension) might have biased the participants' responses during the Conceptualisation Assessment Phase. For instance, they might have revised their initial preferences by answering that whichever references allowed them to successfully complete trials during the Matching Phase matched their conceptualisations best. The pattern of results reported in Table 2 being so similar to that obtained in Experiment 1 (during which

the participants did not perform a matching task) seems to go against this idea. An additional analysis was conducted to confirm that the participants' responses were not significantly different across both experiments. No differences were found (all F s < 1). This point is addressed further in the General Discussion.

Main analysis: Influence of generation and conceptual match on reference reuse during the Matching Phase. The data were analysed following the same rationale as in Experiment 1, except that the dependent variable was reuse during the Matching Phase rather than recall. The model used to analyse the data included Generation (self, other) and Conceptualisation (match, mismatch) as fixed effects. The interaction between these two factors was removed from the model, as it failed to reach statistical significance, $F(1, 139) = 0.01, p = .914$. The outcome variable was the likelihood of reusing a content word during the Matching Phase. The random effect structure included (1) by-dyad random slopes corresponding to Generation and Conceptualisation; (2) by-participant random slopes corresponding to Generation and Conceptualisation; (3) by-item random slopes corresponding to Generation and Conceptualisation. All other random effects (i.e., by-dyad random intercepts, by-participant random intercepts and by-item random intercepts) were removed from the analysis because they caused convergence failure. Removing these random effects had no influence on the outcome of the analysis.

The results are reported in Figure 7. Conceptualisation significantly predicted reuse, $F(1, 29) = 114.57, p < .001$. The participants were more likely to reuse content words which matched their conceptualisation than content words which did not $OR = 15.82, CI_{.95} = 9.34, 26.82$. However, the effect of Generation on reuse failed to reach statistical significance, $F(1, 30) = 2.90, p = .099$. The model parameters are reported in Appendix A (Table A2).

Insert Figure 7 around here

An additional statistical analysis confirmed that the same pattern of results was found even when Acceptance and Short-term reuse were taken into account (Appendix B).

Discussion

The purpose of Experiment 2 was to examine how generation and conceptualisation affect reference reuse in order to offer a better understanding of dialogic decisions about how to refer (Knutsen & Le Bigot, 2014; Knutsen et al., in press). The results confirmed that reference reuse during the Matching Phase was guided by conceptualisation: directors were more likely to reuse a reference when it matched their conceptualisation than when it did not. This is in line with the idea that the reuse of references from the common ground might reflect not only their memory of what was said previously, but also the speakers' tendency to perceive referents in the same way each time they are exposed to them. This replicates Duff et al.'s (2006) work and extends their findings to typical individuals. Unexpectedly, however, the effect of generation on reuse failed to reach statistical significance in this experiment, preventing us from concluding that references which are more readily accessible in memory (as shown in Experiment 1) are also more likely to be reused. This is incompatible with the idea that reference reuse in dialogue is systematically subject to a *self-presentation bias*, as it implies that the conceptualisation match effect sometimes caused participants to favour the reuse of partner-presented references. Importantly, we cannot discard the possibility that the lack of a significant generation effect could be due at least in part to directors trying to adapt their descriptions by favouring the reuse of references which had originally been generated by their partner (i.e., the matcher). The data do not support the idea that this kind of behaviour was systematic across participants and/or across items, but it might have been frequent enough to “cancel” any potential self-generation bias in this study.

General Discussion

The results of Experiment 1 and 2 contribute to a better understanding of the low-level memory and conceptual processes at play during dialogue. First, the results of Experiment 1 suggest that at least some decisions about how to refer (namely, decisions which match one's conceptualisations) are subject to a generation effect: self-generated content words are remembered better than partner-generated ones (Burnett & Bodner, 2014; Rosner et al., 2013; Slamecka & Graf, 1978). This has important theoretical consequences. Indeed, one of the central ideas developed within the framework of the collaborative approach to dialogue is that all partners involved in an interaction come to share a similar representation of their common ground (Clark, 1996; Clark & Marshall, 1981). The current results nuance this assumption by suggesting that dialogue partners may share a representation of their common ground, but that they are subject to a bias in storing and/or retrieving common ground information. In particular, the generation effect implies that references which are readily accessible to A are not necessarily readily accessible to B, and vice-versa.

We have previously suggested that this pattern of results (i.e., each speaker remembering the information they presented better than the information presented by others after the end of an interaction) has direct implications for partner-adaptation in subsequent interactions (Knutsen & Le Bigot, 2015). Indeed, part of the common ground consists in the information mentioned during past interactions between the same dialogue partners (Brennan & Clark, 1996; Clark & Marshall, 1981; Metzging & Brennan, 2003). Such bias in conversational memory could therefore lead each speaker to favour the reuse of self-presented references for subsequent adaptation purposes. However, the results of Experiment 2 do not support this suggestion. Although self-generated references were more readily accessible in memory than partner-generated ones, reference reuse in this experiment was mainly guided by conceptualisation. Such conceptual match effect affected both self- and

partner-generated references. This implies that in a situation where Speaker A hears Speaker B generating a reference that matches his or her (i.e., A's) own perception of the topic under discussion, A is capable of reusing this information later during the interaction – despite it being less readily accessible in memory than self-generated information.

A potential confound in this study was addressed by comparing the participants' responses during the Conceptualisation Assessment Phase in Experiment 1 and Experiment 2. The main concern here was that having the opportunity to discuss the figures again in Experiment 2 (i.e., during the Matching Phase) might have caused the participants to reconceptualise at least some of the stimuli used. For instance, at the end of the Dialogue Phase, Participant X might have believed that the reference "the cat" was most appropriate to refer to one of the figures, but realising that Participant Y had difficulty understanding this reference during the Matching Phase or hearing Participant Y using the alternative reference (e.g., "the fox") during this phase might have altered Participant X's response during the Conceptualisation Assessment Phase. However, the lack of a significant difference in the participants' responses in Experiment 1 (where the tangram figures were not discussed again after the Dialogue Phase) and Experiment 2 (where they were discussed again) is not in support of this possibility. However, another possibility which cannot be discarded is that the participants' conceptualisations might have changed *during* the Dialogue Phase. For instance, when shown a figure during the Dialogue Phase, Participant X might have conceptualised it as a cat, thus leading him or her to generate the reference "the cat", but subsequently hearing Participant Y generating the reference "the fox" during the same phase might have led Participant X to reconceptualise this referent and to now view it as a fox as well. This would not be a major issue for the current study, as this new conceptualisation (which would have potentially affected Participant X's behaviour in phase 2) would have been captured during the Conceptualisation Assessment Phase (i.e., Participant X would have selected "the fox"

rather than “the cat” during this phase). Nonetheless, it would be interesting to examine whether initial conceptualisations and re-conceptualisations affect reference reuse in the same way. The current study was not designed to address this question, but this point should be addressed in future research.

In sum, the main implication of this study is that the notion of “self-presentation bias” should be used with caution, as it seems that that such bias could reflect conceptual match in addition to, or instead of, self-generation. As suggested in the Introduction, it seems reasonable to assume that these two determinants are often confounded in everyday conversation (and potentially in many laboratory experiments on spontaneous dialogue): speakers tend to generate references which match their conceptualisation at the time of initial presentation. However, in cases where a self-presented reference does not match one’s conceptualisation of the referent under discussion, reuse depends more on conceptual match than on generation. Thus, the notions of “self-generation bias” and “conceptual match bias” should be preferred to the single notion of “self-presentation bias” in order to account more precisely for reuse biases in dialogue.

Links with Theoretical Models of Language Production

The results from the current study as well as previous findings (Knutsen & Le Bigot, 2014; Knutsen et al., in press) suggest that memory processes and conceptual processes affect dialogic reference reuse. The idea that different kinds of constraints may affect language production and comprehension is not new in dialogue research (see Hanna, Tanenhaus, & Trueswell, 2003; Kronmüller & Barr, 2015). Precisely, many of the studies focusing on this idea have sought to investigate the time course by which these constraints influence language processing. The current study was not designed to answer the question of *when* accessibility

in memory and conceptual match affect reference production (and whether or not they do so simultaneously), but suggestions can nonetheless be formulated at this point.

Most theoretical models of language production involve a conceptual preparation phase (during which the information regarding which ideas a speaker intends to express and which perspective he or she will use to express these is retrieved) followed by a lexical selection phase (during which the corresponding lexical representations are retrieved) (e.g., Dell, 1986; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Our suggestion is that the conceptual preparation phase is subject to self-generation bias as well as a conceptual match bias. Indeed, decisions to refer *and* decisions about how to refer are made during this phase. Decisions to refer were not investigated in the current study, but there is evidence that such decisions are subject to a self-generation bias (Knutsen et al., in press; see also Knutsen & Le Bigot, 2014) which makes speakers more likely to re-refer to referents corresponding to self-presented references than to re-refer to referents corresponding to partner-presented references.

As for the conceptual match bias, the “old piece of furniture/beautiful wooden table” example discussed in the Introduction can be used to illustrate our point. In this example, Speaker A refers to a table as “an old piece of furniture” and Speaker B refers to the same table as “a beautiful wooden table” (in this example, these two references reflect each participant’s conceptualisation of the referent). Later during the interaction, Speaker A intends to refer to the same table again. To do this, A must first go through the conceptual preparation phase. At this point, A has the choice between at least two different conceptualisations (i.e., the two conceptualisations corresponding to the two references); the conceptual match effect should cause him or her to select the “old piece of furniture” conceptualisation. The activation then spreads to the corresponding node at the lexical level, eventually leading Speaker A to reuse the reference “the old piece of furniture” again.

It is noteworthy that in this example, the preferred conceptualisation is associated with self-generated words. However, imagine a situation in which this is not the case. For instance, both Speaker A and Speaker B conceptualise the table as an old piece of furniture, and B presents the corresponding reference first. Later during the interaction, A intends to refer to the same table again. In this kind of situation, activation should spread from the “old piece of furniture” conceptualisation to the corresponding reference, thus facilitating the production of an initially partner-generated reference. This could also help explain why the generation effect did not significantly affect reuse in Experiment 2: the activation of preferred conceptualisations might have led to the activation of the corresponding references regardless of who had initially generated them.

Although conceptual match plays an important role in dialogic reference production, there are nonetheless a number of situations in which speakers are capable of producing references which do not match their own conceptualisations. This is especially apparent in studies in which experts interact with novices. For instance, computer experts are capable of adapting their speech depending on the level of knowledge of the partners they interact with (Nückles et al., 2006). In such situations, reference production mainly reflects the speakers’ beliefs regarding their partners’ level of knowledge rather than their own conceptualisations.

In sum, the conceptual preparation phase appears to be subject to two main biases: a self-generation bias which affects which referents speakers decide to talk about, and a conceptual match bias which affects how people talk about these referents. Moreover, as mentioned above, the conceptual assessment phase is followed by lexical selection, raising the question of whether and how these two biases affect this phase as well. One possibility is that in cases where a single conceptualisation is associated with both self- and partner-generated words (e.g., A says [about a tangram figure]: “it looks like a guy leaning against a tree” and B replies: “yes, he looks very happy”), a self-generation effect would cause each

speaker to favour the reuse of initially self-generated words. The current study was not designed to address this possibility, although self-generated words being more readily accessible in memory in Experiment 1 would be consistent with this account.

Limitations and Directions for Future Research

One main limitation of the current study is that pairs of participants were explicitly required to generate two different references for each tangram figure discussed. The purpose of this was to distinguish the effects of generation and/or conceptual match on reference reuse; however, it is unreasonable to assume that dialogue partners would spontaneously adopt this kind of behaviour in real-life dialogue. Rather, in such situations, one of the partner might generate a reference matching his or her conceptualisation of the referent under discussion; this reference would then be accepted by the other participant (Clark & Schaefer, 1989). This raises the question of whether accepting a reference necessarily involves adopting the same conceptualisation as the partner who initially presented the reference. Another question raised concerns whether conceptualisations must necessarily belong to the common ground in order for them to be used as cues to reference production. Indeed, in Experiment 2, the reference which matched the director's conceptualisation best systematically belonged to the participants' common ground (as it had been discussed during the Dialogue Phase), implying that the matcher was presumably capable of understanding it (regardless of whether it matched his or her own conceptualisation as well or not). It was therefore not possible to determine whether the directors favoured the reuse of such references because they knew that their partners were capable of understanding them, or whether they simply favoured the reuse of references which were easy to produce for them. Finally, future studies should seek to generalise the current findings to situations in which dialogue partners refer to stimuli other than abstract tangram figures. One of the main features of these pictures is that they are not

usually associated with a predefined label, involving that dialogue partners must explicitly negotiate which references should be used to refer to them (e.g., Clark & Wilkes-Gibbs, 1986; Duff et al., 2006). However, in real-life dialogue, speakers also refer to more consensual referents (e.g., place names, people's names and object names). The influence of conceptual match on the production of such references might be smaller than in the current study, as there might be less room for speakers to think about how they might conceptualise the referents in this kind of situation.

Conclusion

At least two different constraints affect the reuse of references that belong to the common ground: accessibility in memory (which may depend on initial reference generation) and conceptualisation. This finding is compatible with the idea that low-level, “ordinary” processes (i.e., processes which are not specific to dialogue) influence reference production (and comprehension) during dialogue (e.g., Gorman et al., 2013; Horton & Gerrig, 2002, 2005a, 2005b; Horton & Slaten, 2012). This idea was initially developed within the framework on the memory-based approach to dialogue, which emphasizes the role played by automatic memory processes in partner-adaptation. The current findings advocate for a broadening of the scope of this approach, as it seems that reference production depends on conceptual processes as well.

References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed). New York, NY: John Wiley & Sons.
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195–225. http://doi.org/10.1207/s15516709cog2702_3
- Bard, E. G., Hill, R. L., Foster, M. E., & Arai, M. (2014). Tuning accessibility of referring expressions in situated dialogue. *Language and Cognitive Processes*, 29, 928–949. <http://doi.org/10.1080/23273798.2014.895845>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493. <http://doi.org/10.1037/0278-7393.22.6.1482>
- Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, 27, 62–89. <http://doi.org/10.1080/01690965.2010.543363>
- Burnett, A. N., & Bodner, G. E. (2014). Learnin' 'bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences. *Journal of Memory and Language*, 75, 1–13. <http://doi.org/10.1016/j.jml.2014.04.005>
- Clark, H. H. (1996). *Using language*. Cambridge, MA: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

- Clark, H. H., & Marshall, C. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (Vol. 2, pp. 57–63). New York, NY: Association for Computing Machinery.
- Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (Vol. 2, pp. 10–63). Cambridge, MA: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294. http://doi.org/10.1207/s15516709cog1302_7
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39. [http://doi.org/10.1016/0010-0277\(86\)90010-7](http://doi.org/10.1016/0010-0277(86)90010-7)
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. <http://doi.org/10.1037/0033-295X.93.3.283>
- Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, *9*, 140–146. <http://doi.org/10.1038/nn1601>
- Fox Tree, J. E., & Clark, N. B. (2013). Communicative effectiveness of written versus spoken feedback. *Discourse Processes*, *50*(5), 339–359. <http://doi.org/10.1080/0163853X.2013.797241>
- Fukumura, K. (2015). Interface of linguistic and visual information during audience design. *Cognitive Science*, *39*, 1419–1433. <http://doi.org/10.1111/cogs.12207>
- Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What’s learned together stays together: Speakers’ choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 843–853. <http://doi.org/10.1037/a0029467>

- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61. [http://doi.org/10.1016/S0749-596X\(03\)00022-6](http://doi.org/10.1016/S0749-596X(03)00022-6)
- Hjelmquist, E. (1984). Memory for conversations. *Discourse Processes*, *7*, 321–336. <http://doi.org/10.1080/01638538409544595>
- Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*, 1114–1139.
- Horton, W. S. (2008). A memory-based approach to common ground and audience design. In I. Kecskes (Ed.), *Intention, common ground, and the egocentric speaker-hearer* (pp. 189–222). Berlin, Germany/New York, NY: Mouton de Gruyter.
- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589–606. [http://doi.org/10.1016/S0749-596X\(02\)00019-0](http://doi.org/10.1016/S0749-596X(02)00019-0)
- Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 1–35. http://doi.org/10.1207/s15326950dp4001_1
- Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142. <http://doi.org/10.1016/j.cognition.2004.07.001>
- Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition*, *40*, 113–126. <http://doi.org/10.3758/s13421-011-0135-7>
- Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British*

Journal of Psychology, 82(4), 449–462. <http://doi.org/10.1111/j.2044-8295.1991.tb02412.x>

Isaacs, E. A., & Clark, H. H. (1987). Reference in conversations between experts and novices. *Journal of Experimental Psychology: General*, 116, 26–37. <http://doi.org/10.1037/0096-3445.116.1.26>

Jaccard, J. (2001). *Interaction effects in logistic regression*. Thousand Oaks, CA: Sage.

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. <http://doi.org/10.1016/j.jml.2007.11.007>

Jarvella, R. J., & Collas, J. G. (1974). Memory for the intentions of sentences. *Memory & Cognition*, 2, 185–188. <http://doi.org/10.3758/BF03197513>

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). The analysis of repeated measurements: A comparison of mixed-model satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in Statistics - Theory and Methods*, 28, 2967–2999. <http://doi.org/10.1080/03610929908832460>

Kiernan, K., Tao, J., & Gibbs, P. (2012). Tips and strategies for mixed modelling with SAS/STAT procedures. Presented at the 2012 SAS Global Forum, Orlando, FL.

Knutsen, D., & Le Bigot, L. (2012). Managing dialogue: How information availability affects collaborative reference production. *Journal of Memory and Language*, 67, 326–341. <http://doi.org/10.1016/j.jml.2012.06.001>

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, 21, 1590–1599. <http://doi.org/10.3758/s13423-014-0620-7>

- Knutsen, D., & Le Bigot, L. (2015). The influence of reference acceptance and reuse on conversational memory traces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 574–585. <http://doi.org/10.1037/xlm0000036>
- Knutsen, D., Ros, C., & Le Bigot, L. (in press). Generating references in naturalistic face-to-face and phone mediated dialogue settings. *Topics in Cognitive Science*.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343–346. <http://doi.org/10.1037/h0023705>
- Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, *83*, 1–19. <http://doi.org/10.1016/j.jml.2015.03.008>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in language production. *Behavioral and Brain Sciences*, *22*, 1–75.
- Lysander, K., & Horton, W. S. (2012). Conversational Grounding in Younger and Older Adults: The Effect of Partner Visibility and Referent Abstractness in Task-Oriented Dialogue. *Discourse Processes*, *49*(1), 29–60. <http://doi.org/10.1080/0163853X.2011.625547>
- MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197–1202. <http://doi.org/10.3758/s13423-011-0168-8>
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685. <http://doi.org/10.1037/a0018785>

- McMahon, J. M., Pouget, E. R., & Tortu, S. (2006). A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NLMIXED. *Computational Statistics & Data Analysis*, *50*, 3663–3680. <http://doi.org/10.1016/j.csda.2005.08.008>
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201–213. [http://doi.org/10.1016/S0749-596X\(03\)00028-7](http://doi.org/10.1016/S0749-596X(03)00028-7)
- Nückles, M., Winter, A., Wittwer, J., Herbert, M., & Hübner, S. (2006). How do experts adapt their explanations to a layperson's knowledge in asynchronous communication? An experimental study. *User Modeling and User-Adapted Interaction*, *16*, 87–127. <http://doi.org/10.1007/s11257-006-9000-y>
- Rosner, Z. A., Elman, J. A., & Shimamura, A. P. (2013). The generation effect: Activating broad neural circuits during memory encoding. *Cortex*, *49*, 1901–1909. <http://doi.org/10.1016/j.cortex.2012.09.009>
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110–114. <http://doi.org/10.2307/3002019>
- Slamecka, N., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604. <http://doi.org/10.1037/0278-7393.4.6.592>
- Spiers, H. J., Maguire, E. A., & Burgess, N. (2001). Hippocampal amnesia. *Neurocase*, *7*, 357–382. <http://doi.org/10.1076/neur.7.5.357.16245>
- Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics* (Vol. 9, pp. 315–332). New York, NY: Academic Press.
- Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 919–937. <http://doi.org/10.1037/a0036161>

Appendix A: Model parameters

Table A1

Experiment 1 – Model parameters for the main analysis

Covariance parameter estimates						
Parameter	Subject	Estimate	Standard Error			
Intercept	Dyad	0.04	0.07			
Conceptualisation	Dyad	0.12	0.07			
Intercept	Participant	0.02	0.04			
Generation	Participant	0.09	0.05			
Intercept	Item	0.14	0.06			
Conceptualisation	Item	0.07	0.05			

Fixed effects						
Effect	Estimate	Standard Error	DF	<i>t</i> Value	<i>p</i> value	
Intercept	-1.40	0.13	47	-11.10	< .001	
Generation: self	-0.05	0.10	45	-0.50	0.62	
Generation: partner	0					
Conceptualisation: match	2.39	0.18	42	13.51	< .001	
Conceptualisation: no match	0					
Self * Match	0.82	0.18	5793	4.43	< .001	
Self * No match	0					
Partner * Match	0					
Partner * No match	0					

Note. An estimate of 0 implies that this level was used as the baseline in the analysis. Note that in this table, the Generation effect seems to be non-significant, although it was found to be significant in the main analysis ($F(1, 67) = 9.59, p = .003$). This is because in models which include more than one factor, model parameters for one factor are calculated in the *baseline* level of the other factor. In other words, here, the model parameters for Generation correspond to the models parameters for this factor *in the no-match condition only*. The fact that the overall Generation effect was significant and the significant interaction confirm that although the effect of Generation was not significant in the no-match condition, it was significant in the matching condition.

Table A2

Experiment 2 – Model parameters for the main analysis

Covariance parameter estimates			
Parameter	Subject	Estimate	Standard Error
Generation	Dyad	0.21	0.12
Conceptualisation	Dyad	0.42	0.18
Generation	Participant	0.03	0.09
Conceptualisation	Participant	0.07	0.10
Generation	Item	0.21	0.09

Conceptualisation	Item	0.14	0.10			
Fixed effects						
Effect		Estimate	Standard Error	DF	<i>t</i> Value	<i>p</i> value
Intercept		-2.57	0.22	49	-11.61	< .001
Generation: self		0.36	0.21	30	1.70	.099
Generation: partner		0				
Conceptualisation: match		2.76	0.26	29	10.70	< .001
Conceptualisation: no match		0				

Note. An estimate of 0 implies that this level was used as the baseline in the analysis.

Appendix B: Additional analyses (Experiments 1 and 2)**Experiment 1**

To examine whether the participants' performance during the Memory Assessment Phase depended on other determinants than Generation and Conceptualisation, the data were recoded for Acceptance and Short-term reuse.

A presented content word was coded as accepted through verbatim repetition (code 1) when the participant who did not perform the presentation repeated it between the moment when its initiator presented it and the moment when he or she produced another content word. All other presented content words were coded as accepted through another mean (code 0).

All occurrences of content word production which did not count as presentations or acceptances through verbatim repetition were coded as reuses. Following previous work on the self-presentation bias, a reference was counted as reused only if the speech turn in which it occurred was preceded by a minimum of two speech turns during which it did not occur (Knutsen & Le Bigot, 2012). Therefore, each reference presented was either coded as reused (code 1; note that this code was used regardless of how many times the reference had actually been reused) or non-reused (code 0). Note that these two variables (Acceptance and Short-term reuse) were coded at the dyadic level: an accepted reference was coded as accepted from both participants' point of view and a reused reference was coded as reused from both participants' point of view. In the final dataset (i.e., the dataset used in the main analysis), the total number of content words coded as accepted through verbatim repetition was 183 (4.77% of all content words presented) and the total number of content words coded as reused was 162 (5.60% of all content words presented).

A logistic mixed model was used to analyse the data. Following the same rationale as in the main analysis, this model included Generation and Conceptualisation as fixed effects

and recall as the outcome variable. The initial model also included all random intercepts and random slopes justified by the design. It also included by-dyad, by-participants and by-item random slopes corresponding to Acceptance and Reuse. This allowed us to remove from the model the variability associated with these two factors.

The random effects causing model convergence failure were then removed from the model, which did not affect the outcome of the analysis (removing the random slopes corresponding to Acceptance and Short-term reuse during this process would imply that the variability associated with these random effects was equal to zero).

The random effects structure of the final model included (1) by-dyad random slopes corresponding to Conceptualisation, Acceptance and Short-term reuse; (2) by-participant random intercepts and by-participant random slopes corresponding to Generation; (3) by-item random slopes corresponding to Conceptualisation, Acceptance and Short-term reuse.

Generation significantly predicted recall, $F(1, 69) = 10.99, p = .002$. The participants were more likely to recall self-generated content words than partner-generated ones, $OR = 1.45, CI_{.95} = 1.16, 1.82$. Conceptualisation also significantly predicted recall, $F(1, 25) = 408.01, p < .001$. The participants were more likely to recall content words which matched their conceptualisations than content words that did not match their conceptualisations, $OR = 16.43, CI_{.95} = 12.35, 21.86$.

Furthermore, the interaction between these two factors was statistically significant, $F(1, 5793) = 18.84, p < .001$. Additional pairwise comparisons (Sequential Bonferroni) were conducted to offer a better understanding of this pattern of results. These comparisons revealed that the difference between self- and partner-generated references was significant when these references matched the participants' conceptualisation ($p < .001$) but not when these references did not match the participants' conceptualisations ($p = 1.00$). The model parameters are reported in Table B1.

This pattern of results is identical to the one reported in the main analysis. These results confirm that the effects of Generation and Conceptualisation remained statistically significant even when Acceptance and Short-term reuse were taken into account in the analysis.

Table B1

Experiment 1 – Model parameters for the additional analysis

Covariance parameter estimates						
Parameter	Subject	Estimate	Standard Error			
Conceptualisation	Dyad	0.09	0.05			
Acceptance	Dyad	0.34	0.15			
Short-term reuse	Dyad	0.38	0.16			
Intercept	Participant	0.03	0.04			
Presentation	Participant	0.08	0.04			
Conceptualisation	Item	0.07	0.05			
Acceptance	Item	0.15	0.08			
Short-term reuse	Item	0.05	0.07			

Fixed effects						
Effect	Estimate	Standard Error	DF	<i>t</i> Value	<i>p</i> value	
Intercept	-1.05	0.19	47	-5.50	< .001	
Generation: self	-0.03	0.10	45	-0.33	0.74	
Generation: partner	0					
Conceptualisation: match	2.39	0.17	55	14.21	< .001	
Conceptualisation: no match	0					
Self * Match	0.81	0.19	5793	4.34	< .001	
Self * No match	0					
Partner * Match	0					
Partner * No match	0					

Note. An estimate of 0 implies that this level was used as the baseline in the analysis.

Experiment 2

The dialogues between the participants during the Dialogue Phase were coded for Acceptance and Short-term reuse. In the final dataset, the total number of content words coded as accepted through verbatim repetition was 102 (3.46% of all content words presented) and the

total number of content words coded as reused was 95 (3.23% of all content words presented).

This additional analysis was conducted following the same rationale as the additional analysis in Experiment 1. The random effects structure of the final model included the same random intercepts and slopes as the main analysis (i.e., by-dyad random slopes corresponding to Generation and Conceptualisation, by-participant random slopes corresponding to Generation and Conceptualisation and by-item random slopes corresponding to Generation and Conceptualisation). In other words, random slopes corresponding to Acceptance and Short-term reuse were not included in the final model because the variability associated with these random effects was equal to zero, implying that it was not necessary to control for these variables in this analysis (the results were identical regardless of whether or not these slopes were included in the model). Because the random structure was identical to that used in the main analysis, the results were also identical to those of the main analysis (i.e., only a significant effect of Conceptualisation was found, $F(1, 29) = 114.57, p < .001$). The model parameters were also identical to those found in the main analysis (see Table A2; because the parameters were identical to those reported previously, they are not reported again in this appendix). This confirms that Conceptualisation significantly affected reuse in this experiment, regardless of potential acceptance and/or short-term reuse.