Running head: Reference reuse in dialogue

Capturing egocentric biases in reference reuse during collaborative dialogue

Dominique KNUTSEN and Ludovic LE BIGOT

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

Corresponding author

Dominique KNUTSEN

Centre de Recherches sur la Cognition et l'Apprentissage – CeRCA – CNRS UMR 7295

MSHS Bât A5

5, rue Théodore Lefebvre

F – 86000 Poitiers – France

Tel: +33 (0)5 49 45 46 82

Fax: +33 (0)5 49 45 46 16

E-mail: dominique.knutsen@univ-poitiers.fr or dom.knutsen@gmail.com

Abstract

Words that are produced aloud – and especially self-produced ones – are remembered better than words that are not, a phenomenon labeled the production effect in the field of memory research. Two experiments were conducted to determine if this effect can be generalized to dialogue, and how it might affect dialogue management. Triads (Experiment 1) or dyads (Experiment 2) of participants interacted to perform a collaborative task. Analyzing reference reuse during the interaction revealed that the participants reused more the references that they had presented themselves on one hand and those that had been accepted through verbatim repetition on the other. Analyzing reference recall suggested that the greater accessibility of self-presented references was only transient. Moreover, among partner-presented references, those discussed while the participant actively took part in the conversation were more likely to be recalled than those discussed while the participant was inactive. These results contribute to a better understanding of how individual memory processes might contribute to collaborative dialogue.

Keywords: dialogue; production effect; referential communication; accessibility in memory; egocentrism

Capturing egocentric biases in reference reuse during collaborative dialogue

Dialogue is a collaborative activity during which speakers interact to reach a common goal, such as establishing a route together (Clark, 1996; Clark & Wilkes-Gibbs, 1986). As speakers interact, they increment their *common ground*, which consists in the information that they are aware of sharing (Clark & Marshall, 1981). The common ground includes the references produced by the speakers earlier in the interaction to refer to objects and entities (e.g., the landmarks to be encountered on a route).

Information is *grounded* (i.e., added to the common ground) through a joint contribution process (Clark & Schaefer, 1989). One of the speakers starts by presenting a piece of information (e.g., a reference); the latter then accepts this information by signaling that it has been understood well enough for current purposes (Clark & Brennan, 1991). Reference acceptance can be explicit, with the addressee accepting the reference through verbatim repetition or through anaphoric repetition (i.e., the addressee uses a pronoun to re-refer to the same object or entity). Acceptance can also be implicit, with the addressee initiating the next relevant speech turn. A grounded reference can potentially be reused by either speaker during the interaction (Brennan & Clark, 1996).

Dialogue being a collaborative activity implies that all speakers put efforts into achieving mutual understanding (Clark, 1996). One way of doing so consists in each speaker using the common ground to determine which references his or her partner is capable of understanding, reasoning that he or she should be capable of understanding a reference that was successfully understood earlier in the interaction (Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987: Haviland & Clark, 1974). However, all the references from the common ground are not equally likely to be reused, as this depends on their accessibility in memory (e.g., Horton, 2008). This is partly in line with an egocentric approach to dialogue, as reference

production depends on the speakers' state of mind rather than on the addressee's (e.g., Barr & Keysar, 2002).

The current study seeks to investigate reference reuse further. A series of studies conducted in the field of memory research has shown that words produced aloud are remembered better than words read silently (Forrin, Ozubko & MacLeod, 2012; MacLeod, Gopie, Neary, & Ozubko, 2010; Ozubko, Hourihan, & MacLeod, 2012); this *production effect* is all the stronger when the production is self-performed (MacLeod, 2011). These findings could have implications for reference reuse in dialogue. First, all speakers involved in dialogue might present references at some point. Self-produced words being more readily accessible suggests that self-presented references should be reused more often than partner-presented ones. Second, acceptance sometimes involves verbatim repetition (Clark & Brennan, 1991). This should cause repeated references to benefit from a self-production effect from the addressee's perspective as well. Furthermore, dialogue partners are exposed to references accepted through verbatim repetition twice (once at the time of presentation and once at the time of acceptance). Because the production effect concerns both self- and partner-produced words, repeated references should benefit from a production effect from each partner's perspective. An additional question concerns whether such accessibility differences persist after the end of an interaction, as dialogue partners might sometimes need to resort to the common ground established during past interactions. Memory accessibility after the end of the interaction might also depend on how many times a reference was actually produced during the interaction.

Moreover, multipartite dialogue involves both *ratified participants* (participants addressing or being addressed by a partner) and *side participants* (participants addressing no one and being addressed by no one) (Clark, 1996). Side participants gather common ground (Wilkes-Gibbs & Clark, 1992), but they do not have the opportunity to actually produce

references. Thus, the accessibility in memory of references produced while one was a side participant should be fairly low. This would be consistent with the idea that indirectly established common ground has a weaker influence on reference production (Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013), and with the more general idea that active learning is more efficient than passive learning (e.g., Bonwell & Sutherland, 1996).

Two experiments were conducted to address these assumptions. Participants referred to landmarks as they performed a route description task. In Experiment 1, triadic dialogue was investigated to determine whether the conditions in which a reference is initially grounded affect its subsequent reuse. Reference recall was then used to assess reference accessibility in memory after the end of the interaction. Experiment 2 sought to replicate the findings obtained in Experiment 1 in dyadic dialogue.

Experiment 1

Participants

Fifty-four native French speaking students signed an informed consent form before taking part in the experiment for course credit.

Material and Apparatus

Three identical versions of a map featuring 25 monuments, nine squares, 57 streets names and three points (A, B and C; Figure 1a) and three identical blank versions of this map (Figure 1b) were printed. The interactions were recorded using two microphones and a digital recorder.

Figure 1. Maps used during Experiment 1: (a) map with landmarks used during the dialogue phase (left panel); (b) blank map used during the drafting phase (right panel).

Task and procedure

Three participants took part in the experiment. Each sat facing a different wall of the experimental room so that they could not communicate through nonlinguistic cues. Their task was to establish a touristic route for a person who had no prior knowledge of the town represented on the maps. They knew that they would have to individually write out the entire route after the interaction but they did not know that they would not have access to the initial map while doing so.

During the first part of the experiment (dialogue phase), the three participants used their maps to agree on a route running from A to B, from B to C and to C back to A. Each section was discussed by two ratified participants only, with the third, side participant listening but not being able to intervene. For instance, P01 and P02 would discuss section A-B, P02 and P03 would discuss section B-C and P01 and P03 would discuss section C-A (the three participants were identified depending on their random entry order in the room). This phase lasted for a maximum of 20 minutes. The time spent on each route section was not predetermined in advance.

During the second phase of the experiment (drafting phase), the participants had a maximum of fifteen minutes to individually write out the entire route. They were given a

blank version of the map to guide their recall. They could not communicate nor write anything on their map during this phase.

Experimental design

Three within-participants IVs were used. The first one was reference status. From each participant's point of view, a reference had either been self-presented or presented by a partner while one was a ratified participant or a side participant.

The second IV was acceptance type. Within each triad, a reference had either been accepted through verbatim repetition, implicitly or anaphorically.

The third IV was the number of times a reference was reused within the triad. This was a standardized continuous IV.

Data Coding

The interactions between the participants were transcribed and coded for presentation, acceptance and reuse (see the Appendix for more detail).

When a reference was presented for the first time in a triad, it was coded as presented for P01, P02 and P03. Who the current speaker and the ratified participants were at the time was used to code for reference status from each participant's point of view. The evidence produced by the other ratified participant between the moment when the reference was presented and the moment when the initiator of the reference produced another reference was examined to code the reference for acceptance type. All other occurrences of reference production were classified as reuse; the only criterion was that reuse needed to occur in a speech turn preceded by a minimum of two speech turns during which the reference was not produced, which helped distinguish reuse from simple repetition. Two different levels of coding were used: at the participant level, reuse was coded as a dichotomous variable (a

reference could be reused by a participant or not, regardless of how many times he or she reused it), whereas it was coded as a continuous variable at the triadic level (this coding reflected how many times each reference was reused in the triad, regardless of who had reused it).

Data coding – drafting phase. The routes wrote out by the participants were transcribed and coded for reference recall: for each participant, each presented reference could either be recalled or not.

Results and discussion

During the dialogue phase, the average number of words produced per triad was 1318.94 (*SD* = 704.18) (A-B: 428.17 (*SD* = 192.40), B-C: 463.28 (*SD* = 333.58); C-A: 427.50 (*SD* = 274.58). A total of 734 references were presented (40.78 (*SD* = 7.76) per triad on average and 13.59 (*SD* = 5.04) per participant on average). Among these, 148 (20.16%) were accepted through verbatim repetition, 355 (48.37%) were accepted anaphorically and 231 (31.47%) were accepted implicitly. During the drafting phase, the mean number of words per individual route was 181.24 (*SD* = 81.94) and the average number of references recalled was 8.61 (*SD* = 3.77).

The data were analyzed in SPSS 22.0. Multilevel models were used to account for the nesting of the participants within the triads. Such models include random intercepts to account for variability across participants (and potentially items) and random slopes to account for participants' (and items') different sensitivity to independent variables. Whenever possible, all random effects justified by the experimental design should be included (Barr, Levy, Scheepers, & Tily, 2013). The analyses reported hereafter included (1) by-triad and by-participant intercepts to account for variability across triads and across participants, (2) by-participant random slopes corresponding to the IVs to account for the participants' different

sensitivity to these variables, (3) by-participant random slopes corresponding to the landmarks to account for the participants potentially behaving differently depending on the landmarks. An additional factor, section responsibility, identified which sections (AB, BC and CA) had been discussed by which participants. By-participant random slopes corresponding to this factor were included to account for its potential influence on reference production. An identity variance-covariance matrix was used. The Satterthwaite correction was applied to estimate the degrees of freedom in the analyses.

Because the two variables used as outcomes in the analyses were binary, logistic mixed models were used. One parameter returned by logistic regression models is the odds ratio (*OR*; Jaccard, 2001), which is informative with regard to the effect size (see Agresti, 2002). Only the significant effects were included in the models. When necessary, additional comparisons were conducted using paired comparisons (Sequential Bonferroni, $p < .05$).

Reference reuse – dialogue phase. A total of 809 references were reused at least once, but only the data corresponding to the references that had been presented – from each participant's point of view – while one was a ratified participant were considered, as the participants seldom reused the references presented while they were side participants. The 43 cases where this happened were discarded from further analysis.

The model included reference status and acceptance type as fixed effects and whether the reference was reused as the outcome variable (Table 1).

Table 1

Experiment 1 – Number of references reused during the dialogue phase as a function of reference status and acceptance type

	Presented by self	Presented by other	**Total**
Accepted through verbatim repetition	92 (.62)	89 (.60)	**181 (.61)**
Accepted anaphorically	205 (.58)	182 (.51)	**387 (.55)**
Accepted implicitly	111 (.48)	87 (.38)	**198 (.43)**

Total	**408 (.56)**	**358 (.49)**	

Note: The proportions reported in brackets were obtained by dividing the number of references reused in each category by the total number of references initially presented in that category, which can be found at the beginning of the Results section. (Reminder: among the references presented, 148 were accepted through verbatim repetition, 355 were accepted anaphorically and 231 were accepted implicitly. For instance, to obtain the proportion .62 for self-presented references accepted through verbatim repetition, 92 was divided by 148.)

Reference status significantly predicted reference reuse, $F(1, 386) = 5.71$, $p = .017$. Self-presented references were more likely to be reused than partner-presented ones, $OR = 1.36$, $CI_{.95} = 1.06, 1.75$. Acceptance type also significantly predicted reference reuse, $F(2, 862) = 5.51$, $p = .004$. References accepted anaphorically and implicitly were less likely to be reused than those accepted through verbatim repetition, $OR = .67$, $CI_{.95} = .49, .93$, $p = .017$, and $OR = .55$, $CI_{.95} = .39, .79$, $p = .001$. References accepted anaphorically were more likely to be reused than references accepted implicitly (Sequential Bonferroni, $p < .05$). This confirms that reference reuse depends on how a reference was initially grounded.

Reference recall – drafting phase. The participants recalled a total of 480 references. Among these, 15 that had not been presented during the dialogue phase were discarded from the analysis. The model included reference status, acceptance type and the number of times a reference had been reused by the triad as fixed effects and whether the reference was recalled as the outcome variable (Table 2). Note that although the number of reuses was measured at the triad level, this analysis focused on recall at the participant level.

Table 2

Experiment 1 – Number of references recalled during the drafting phase as a function of reference status and acceptance type

	Presented by self	Presented by other while one was a ratified participant	Presented by other while one was a side participant	**Total**
Accepted through verbatim repetition	50 (.34)	38 (.26)	33 (.22)	**121 (.27)**

Accepted anaphorically	75 (.21)	78 (.22)	57 (.16)	**184 (.17)**
Accepted implicitly	49 (.21)	47 (.20)	38 (.16)	**160 (.23)**
Total	**174 (.24)**	**163 (.22)**	**128 (.17)**	

Note. The corresponding proportions are reported in brackets. (Reminder: among the references presented, 148 were accepted through verbatim repetition, 355 were accepted anaphorically and 231 were accepted implicitly.)

Reference status significantly predicted reference recall, $F(2, 1741) = 4.76$, $p = .009$. References presented by a partner while one was a side participant were less likely to be recalled than self-presented ones, $OR = .66$, $CI_{.95} = .50, .86$, $p = .003$. However, the difference between the references presented by a partner while one was a ratified participant and self-presented ones failed to reach statistical significance, $p = .442$. The difference between references presented by a partner while one was a ratified participant and references presented while one was a side participant also failed to reach statistical significance (Sequential Bonferroni, $p > .05$).

Acceptance type also significantly predicted reference recall, $F(2, 2196) = 4.03$, $p = .018$. References accepted anaphorically were less likely to be recalled than references accepted through verbatim repetition, $OR = .67$, $CI_{.95} = .50, .88$, $p = .005$. However, the difference between implicitly accepted references and references accepted through verbatim repetition failed to reach statistical significance, $p = .104$. The difference between references accepted anaphorically and references accepted implicitly also failed to reach statistical significance (Sequential Bonferonni, $p > .05$).This pattern of results only partly replicates the pattern obtained for reference reuse.

The number of reuses during the dialogue phase also predicted reference recall, $F(1, 180) = 79.06$, $p < .001$. The odds of recalling a reference increased with the number of reuses in the triad, $OR = 1.72$, $CI_{.95} = 1.53, 1.95$.

To recap, Experiment 1 shows that self-presented references are more likely to be reused during dialogue than partner-presented references, and that references accepted through verbatim repetition are more likely to be reused than references accepted through other means. Some of these effects have a longer-term influence on reference memorization. Experiment 1 also sheds light on the influence of active participation in the dialogue on subsequent reference memory. However, one potential limitation of this experiment is that it involved triads of participants: the results could be due to the ratified participants knowing that a side participant was listening to them and making extra efforts to repeat the references during the interaction. A second experiment was thus conducted to attempt to replicate the results reported above in a dyadic dialogue situation. Dyads of participants performed the same task as in Experiment 1; the only difference was that both participants acted as ratified participants during the entire dialogue phase.

Experiment 2

Participants

Fifty-four students were recruited under the same conditions than in Experiment 1.

Material

The material was similar to that used in Experiment 1, except that the map only featured two points (A and B; Figure 2).

Figure 2. Maps used during Experiment 2; (a) map with landmarks used during the dialogue phase (left panel); (b) blank map used during the drafting phase (right panel).

Task and procedure

The task and procedure were similar to those of Experiment 1. The participants' task was to establish a return route between A and B.

Experiment design and data coding

The experimental design and coding were the same as in Experiment 1, except that the reference status IV only had two modalities (self- and partner-presented).

Results

During the dialogue phase, the average number of words produced per dyad was 1378.33 (*SD* = 756.60). A total of 1002 references were presented (37.11 (*SD* = 11.45) per dyad on average and 18.56 (*SD* = 6.70) per participant on average). Among these, 201 (20.06%) were accepted through verbatim repetition, 506 (50.50%) were accepted anaphorically and 295 (29.44%) were accepted implicitly. During the drafting phase, the average number of words per individual route was 159.70 (*SD* = 59.38) and the average number of references recalled was 8.61 (*SD* = 3.77).

The statistical analyses were conducted in the same way as those reported in Experiment 1, except that the random part of the model included no section responsibility factor.

Reference reuse – dialogue phase. A total of 1032 references were reused at least once (Table 3).

Table 3

Experiment 2 – Number of references reused during the dialogue phase as a function of reference status and acceptance type

	Presented by self	Presented by other	**Total**
Accepted through verbatim repetition	128 (.64)	108 (.54)	**236 (.59)**
Accepted anaphorically	281 (.56)	255 (.50)	**536 (.53)**
Accepted implicitly	141 (.48)	119 (.40)	**260 (.44)**
Total	**550 (.55)**	**482 (.48)**	

Note. The corresponding proportions are reported in brackets. (Reminder: Among the references presented, 201 were accepted through verbatim repetition, 506 were accepted anaphorically and 295 were accepted implicitly.)

Reference status significantly predicted reference reuse, $F(1, 245) = 4.32$, $p = .039$. Self-presented references were more likely to be reused than partner-produced ones, $OR = 1.26$, $CI_{.95} = 1.01, 1.58$. Acceptance type also significantly predicted reference reuse, $F(2, 468) = 7.74$, $p < .001$. References accepted implicitly were less likely to be reused than references accepted through verbatim repetition, $OR = .58$, $CI_{.95} = .43, .79$, $p < .001$. However, the difference between references accepted anaphorically and references accepted through verbatim repetition failed to reach statistical significance. References accepted anaphorically were more likely to be reused than references accepted implicitly (Sequential Bonferroni, $p < .05$). As in Experiment 1, reuse depended on the circumstances in which a reference was initially grounded.

Reference recall – drafting phase. The participants recalled a total of 460 references (Table 4). Among these, 17 that had not been presented during the dialogue phase were discarded from the analysis.

Table 4

Experiment 2 – Number of references recalled during the drafting phase as a function of reference status and acceptance type

	Presented by self	Presented by other	**Total**
Accepted through verbatim repetition	53 (.26)	52 (.26)	**105 (.26)**
Accepted anaphorically	116 (.23)	117 (.23)	**233 (.23)**
Accepted implicitly	55 (.19)	50 (.17)	**105 (.18)**
Total	**224 (.22)**	**219 (.22)**	

Note. The corresponding proportions are reported in brackets. (Reminder: Among the references presented, 201 were accepted through verbatim repetition, 506 were accepted anaphorically and 295 were accepted implicitly.)

The influence of reference status on reference recall failed to reach statistical significance, $F < 1$. As in Experiment 1, the self-presentation benefit was attenuated after the end of the interaction.

Acceptance type significantly predicted reference recall, $F(2, 752) = 3.63$, $p = .027$. References accepted anaphorically and references accepted implicitly were less likely to be recalled than references accepted through verbatim repetition, respectively $OR = .72$, $CI_{.95} = .52, .99$, $p = .043$ and $OR = .62$, $CI_{.95} = .43, .88$, $p = .008$. The difference between references accepted anaphorically and references accepted implicitly failed to reach statistical significance (Sequential Bonferonni, $p > .05$).

The number of reuses during the dialogue phase also predicted reference recall, $F(1, 139) = 104.97$, $p < .001$. The odds of recalling a reference increased with the number of reuses in the dyad, $OR = 2.26$, $CI_{.95} = 1.93, 2.64$.

The overall pattern of results obtained in Experiment 1 was replicated in Experiment 2. The only differences pertained to acceptance (Table 5). Such differences were probably due to the number of references accepted through verbatim repetition, anaphorically and implicitly varying a lot, which might have caused the statistical power of the analyses to decrease. Nonetheless, the general pattern that emerges here is that references accepted through verbatim repetition were reused more and recalled better than other references, which is consistent with the production effect hypothesis.

Table 5

Summary of the results for reference reuse and recall as a function of acceptance type in Experiments 1 and 2

	Experiment 1	Experiment 2
Reuse		
Verbatim repetition vs. anaphoric acceptance	Reuse more likely when repeated verbatim	Failed to reach statistical significance
Verbatim repetition vs. implicit acceptance	Reuse more likely when repeated verbatim	Reuse more likely when repeated verbatim
Anaphoric acceptance vs. implicit acceptance	Reuse more likely when accepted anaphorically	Reuse more likely when accepted anaphorically
Recall		
Verbatim repetition vs. anaphoric acceptance	Recall more likely when repeated verbatim	Recall more likely when repeated verbatim
Verbatim repetition vs. implicit acceptance	Failed to reach statistical significance	Recall more likely when repeated verbatim
Anaphoric acceptance vs. implicit acceptance	Failed to reach statistical significance	Failed to reach statistical significance

General Discussion

This study focused on the influence of grounding on subsequent reference accessibility. In line with MacLeod (2011), the participants in both experiments showed an egocentric bias towards reusing self-presented references more, suggesting that references from the common

ground continue to "belong" to their initiator during the interaction. However, this effect was attenuated after the end of the interaction. One possible explanation stems from the fact that the references that had been reused more were also recalled better: repeated reuse of partner-presented references by their initiator might have caused the accessibility of these initially less accessible references to increase, thus eventually attenuating the difference between self- and partner-presented references. In addition, the production effect was all the stronger as the number of exposures to a reference at the time of grounding increased, as references accepted through verbatim repetition were reused more by all dialogue partners. These references being reused more could also help explain why they were remembered better. Finally, reference accessibility after the end of the interaction depended on the role played during the interaction (in line with Gorman et al., 2013). This could be due to side participants not having the opportunity to produce the references under discussion themselves; it could also be due to references produced by others constituting weaker episodic traces and therefore being remembered less well (e.g., Andersson & Rönnberg, 1997). The reuse of such references was not investigated directly in this study, due to the task constraints. However, in a situation where speakers do have the opportunity to reuse such references, the fact that they are less accessible in memory could have an influence on dialogue management.

These findings help bridge a theoretical gap between research on memory and research on dialogue, showing how "ordinary" memory mechanisms might constrain higher-level processes involved in collaboration (e.g., Horton, 2008). Specifically, these findings build on Clark and Schaefer's (1989) model by showing how accessibility differences that appear during the first two steps of dialogue management, presentation and acceptance, strongly influence the third step – namely reuse. A number of studies have sought to determine *whether* speakers are capable of taking common ground into account during dialogue (e.g., Horton & Keysar, 1996), leading to the idea that common ground and other sources of

information might both constrain language processing during dialogue (e.g., Hanna, Tanenhaus, & Trueswell, 2003). Our findings contribute to this debate by showing that even when common ground is taken into account, each speaker behaves egocentrically, as what is taken for common ground depends on reference accessibility in memory from each speaker's point of view. Thus, a collaborative behavior is not necessarily non-egocentric.

Moreover, this study shows that reference accessibility varies as the interaction unfolds. We have suggested that the self-presentation egocentric bias could be (partly) compensated for through reuse. Indeed, within a dyad (for instance), Speakers A and B's egocentric biases are complementary, as Speaker A's self-presented references correspond to Speaker B's partner-presented references, and vice-versa. Each speaker being more likely to reuse his or her own self-presented references causes all references to be equally likely to be reused at the dyadic level; such repeated exposure could then contribute to decreasing initial individual biases. However, not all differences in accessibility diminish during the interaction. Contrary to the self-presentation benefit, the accessibility of references accepted through verbatim repetition either remains constant or increases as the interaction unfolds, as both speakers are initially equally likely to reuse these.

The self-presentation benefit decreasing as the verbatim repetition benefit increases could help explain why it is generally considered that a reference that belongs to the common ground can be reused by either speaker (e.g., Brennan & Clark, 1996). This study has allowed us to capture a transient egocentric bias whose effects could not be observed in the longer term. While this effect decreases for *each* speaker, the accessibility of references accepted through verbatim repetition either remains constant or increases for *both* speakers, thus causing the same references to become accessible to them. Thus, provided that they are given enough time, the speakers' representations of the common ground become increasingly similar: the references that are most accessible to one speaker are also most accessible to the

other. From a broader perspective, these findings are in line with the idea that episodic memory traces can be modified depending on the re-use that is made of this information (see Marsh, 2007).

Importantly, the rationale concerning bias complementarity holds mainly for *symmetric* tasks such as the one used in the current study. Not all tasks involve symmetric dialogue. For instance, Clark and Wilkes-Gibbs (1986) asked directors to describe tangram figures to matchers so that the latter could rearrange their figures in a predefined order. In such cases, only one of the participants' initial knowledge is necessary to perform the task. Thus, the speakers' biases are not complementary; the speaker who does not have initial access to the relevant information needs to make extra effort to acquire partner-produced information.

In conclusion, these results contribute to a better understanding of how individual memory processes affect the collaborative processes at play during dyadic and triadic dialogue. They show that the accessibility in memory of grounded references depends on *how* these were initially grounded and they also show how accessibility might vary as the interaction unfolds.

Acknowledgments

This work was conducted as part of the first author's PhD and was supported by the Direction Générale de l'Armement and Région Poitou-Charentes.

References

Agresti, A. (2002). *Categorical Data Analysis, Second Edition.* New York: John Wiley & Sons.

Andersson, J., & Rönnberg, J. (1997). Cued memory collaboration: Effects of friendship and type of retrieval cue. *European Journal of Cognitive Psychology, 9,* 273-287. doi: 10.1080/713752558

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46,* 391-418. doi:10.1006/jmla.2001.2815

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. doi:10.1016/j.jml.2012.11.001

Bonwell, C. C., & Sutherland, T. E. (1996). The active learning continuum: Choosing activities to engage students in the classroom. *New Directions for Teaching and Learning, 67,* 3-16. doi:10.1002/tl.37219966704

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. doi:10.1037/0278-7393.22.6.1482

Clark, H. H. (1996). *Using language*. Cambridge: 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: APA Books.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13,* 259-294. doi:10.1016/0364-0213(89)90008-6

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39. doi:10.1016/0010-0277(86)90010-7

Forrin, N. D., Ozubko, J. D., & MacLeod, C. M. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40,* 1046-1055. doi:10.3758/s13421-012-0210-8

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. doi:10.1037/a0029467

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective taking on domains of referential interpretation. *Journal of Memory and Language, 49,* 43-61. doi:10.1016/S0749-596X(03)00022-6

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521. doi:10.1016/S0022-5371(74)80003-4

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/New York: Mouton de Gruyter.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59,* 91-117. doi:10.1016/0010-0277(96)81418-1

Isaacs, I. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116,* 26-37. doi:10.1037/0096-3445.116.1.26

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

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin and Review, 18,* 1197-1202. doi: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. doi:10.1037/a0018785

Marsh, E. J. (2007). Retelling is not the same as recalling: Implications for memory. *Current Directions in Psychological Science, 16,* 16-20. doi: 10.1111/j.1467-8721.2007.00467.x

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory, 20,* 717-727. doi:10.1080/09658211.2012.699070

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31,* 183-194.

Appendix: Dialogue sample and coding example

Triad	Speech turn	Route section	Ratified participants	Side participant	Current speaker	Utterance content (English translation)	Comments	Coding
T03	(31)	B to C	P01 and P03	P02	P03	yeah and after that we could take hm Charles de Fitte alley	The reference "Charles de Fitte alley" is produced for the first time.	This reference is coded as *presented*, which allows identifying the status of this reference from each participant's point of view.
T03	(32)	B to C	P01 and P03	P02	P01	yeah yeah why not after that we can walk back and continue on Charles de Fitte alley	P01 produces the reference "Charles de Fitte alley" before P03 produces another reference.	This reference is coded as *accepted through verbatim repetition.*
	(33) … (47)							
T03	(48)	C to A	P02 and P03	P01	P02	hm it would be nice to go to the covert market […]	The reference "covert market" is produced for the first time.	This reference is coded as *presented* and reference status is identified.
T03	(49)	C to A	P02 and P03	P01	P03	yeah	P03 produces evidence that the reference was understood.	This reference is coded as *accepted anaphorically*.
T03	(50)	C to A	P02 and P03	P01	P02	walk up the street and arrive next to the covert market	The reference "covert market" is produced in a speech turn that is not preceded by two speech turns that do not contain this reference.	This reference is not coded as *reused* by P02.

T03	(51) … (79)							
T03	(80)	B to C	P01 and P03	P02	P03	we walk down Charles de Fitte alley	The reference "Charles de Fitte alley" is produced in a speech turn that is preceded by two speech turns that do not contain this reference.	This reference is coded as *reused* by P03.

Landmark	Triad	Speaker	Presented in the triad	Reference status	Accepted by self	Acceptance type	Reuse by self (1/0)	Number of reuses in triad
Charles de Fitte alley	T03	P01	1	Other / ratified participant	Yes	Repetition	0	1
	T03	P02	1	Other / side participant	No	Repetition	0	1
	T03	P03	1	Self	No	Repetition	1	1
Covert market	T03	P01	1	Other / side participant	No	Anaphoric	0	0
	T03	P02	1	Self	No	Anaphoric	0	0
	T03	P03	1	Other / ratified participant	Yes	Anaphoric	0	0

Note. To illustrate the coding scheme used, consider the presentation, acceptance and reuse of the reference "Charles de Fitte alley". This reference is produced for the first time during speech turn 31. At the time of presentation, P03 is identified as the current speaker, P01 is identified as the other ratified participant and P02 is identified as the side participant. For all three participants, this reference is coded as presented in the triad; because the basic analysis unit used in this study was the participant and not the triad, each presented reference was included once per participant in the dataset. Reference status is then identified from each participant's point of view: for P01, this reference is coded as presented by a partner while he or she was a ratified participant; for P02, it is coded as presented by a partner while he or she was a side participant; for P03, it is coded as presented by self. The reference is then accepted by the other ratified participant, P01. In this example, the reference "Charles de Fitte alley" is repeated verbatim by P01 before P03 produces another reference: it is thus coded as repeated verbatim. This reference could have been accepted anaphorically, in which case P01 would have used a pronoun to re-refer to the same referent (e.g., "I can see *it* on my map") or would have said "yes" or "okay" (e.g., implying "okay [we could take *this alley*]"; this is the case further on in the interaction, when P02 presents the reference "the covert market" and that P03 accepts it by saying "yes"). This reference could also have been accepted implicitly, in which case P01 would have simply initiated the next relevant speech turn (e.g., P03: "we could take hm Charles de Fitte alley"; P01: "the next stop is the museum"). Finally, at the end of the example, the reference "Charles de Fitte alley" is reused by P03. At the participant

level, it is thus coded as reused for P03 and as non-reused by P01 and P02. At the triadic level, it is coded as reused once by the triad. The purpose of the coding example is to present how these different coding levels were represented in the datasheet used to perform the analyses.