Entrainment on the Move and in the Lab: The Walking Around Corpus

Susan E. Brennan (susan.brennan@stonybrook.edu)
Department of Psychology, Stony Brook University,
Stony Brook, NY 11794-2500 USA

Katharina S. Schuhmann (katharina.schuhmann@stonybrook.edu)
Department of Linguistics, Stony Brook University,
Stony Brook, NY 11794-4376 USA

Karla M. Batres (karla.batres@stonybrook.edu)
Department of Psychology, Stony Brook University,
Stony Brook, NY 11794-2500 USA

Abstract

We examined lexical choice and variability in referring expressions during direction-giving to pedestrians. The Walking Around Corpus comprises an experimentally parameterized collection of spontaneous spoken dialogues produced by 36 pairs of people communicating by mobile telephone; it provides both a testbed for lexical entrainment "in the wild" as well as a resource for pedestrian navigation applications. A stationary partner (the Giver) directed a mobile partner (the Follower) to walk about 1.8 miles, to 18 destinations on a medium-sized campus. Givers viewed a map marked with target destinations, labels, and photos. Followers carried a cell phone with GPS and a digital camera in order to photograph the destinations they visited; Givers monitored Followers' progress as a cursor on a map display. Immediately after the navigation task, Followers returned to the lab and both were tested individually on their spatial ability and memory for the destinations. Next, the Experimenter attempted to interfere with any conceptual precedents they had established by giving Followers printed copies of the photos they had just taken and prompting them (sometimes with competing labels) to identify each destination. Finally, each pair participated in 6 rounds of a more traditional referential communication lab task to repeatedly match duplicate copies of the Follower's pictures of the destinations. Results include significant rates of lexical entrainment, evidence for partner-specific conceptual pacts, and that joint navigation efficiency is affected by direction-givers' spatial ability. The Walking Around Corpus is available to the research community.

Keywords: Referential communication; conceptual pacts; entrainment; collaboration; mobile communication; GPS applications; pedestrian navigation; spatial cognition.

Introduction

Speakers make many expressive choices in conversation, leading to enormous variability in spontaneous speech. These choices emerge not only from their individual proclivities and the availability of words in memory, but also from collaboration with conversational partners (Clark & Wilkes-Gibbs, 1986). There is much less variation *within* than *between* conversations; speakers are more likely to continue using the same term with the same addressee than with a new addressee (Brennan & Clark, 1996). Studies of referential communication in laboratory settings show that lexical choice is influenced not only by precedent (what a speaker has said previously) and strength of precedent (e.g., how often a referent has been discussed), but also by partner (whether the precedent was established with the current partner or a different one) but that it can be maintained via interactive cues provided during grounding (ibid).

We have proposed that *lexical entrainment* marks two speakers' beliefs that they are referring to the same thing; in fact, *breaking a conceptual pact*, such as hearing a speaker inexplicably abandon an entrained-upon expression for a new expression when referring to the same referent, results in slower comprehension than hearing the same new expression from a different (new) partner (Metzing & Brennan, 2003; Matthews, Lieven, & Tomasello, 2010). A conceptual pact does not reflect a rigid mapping between expression and referent, but emerges instead from a flexible, temporary agreement arrived at with an interactive partner to take a particular perspective on the referent; as such, it is highly dependent on context. Although there is evidence that the partner associated with a conceptual pact is represented by a cue in memory (Horton, 2007; Horton & Gerrig, 2005), this does *not* pre-empt an expression and referent from being easily associated with a different partner (Barr & Keysar, 2002; Brennan & Clark, 1996), nor does having an existing conceptual pact with a partner inhibit a new expression-referent association from being encoded with that partner when the pragmatic context changes (ibid).

Lexical entrainment has been demonstrated in adults (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1986) and in children (Matthews et al., 2010); in experts and in novices (Isaacs & Clark, 1987); in native and native/non-native conversations (Bortfeld & Brennan, 1997); and for lexicalized and non-lexicalized referents (Gergle, Kraut, & Fussell, 2013; Schober & Carstensen, 2010). It has also been measured in statistical analyses of the transcripts of speech corpora (e.g., Nenkova, Gravano, & Hirschberg, 2008's studies of the *Switchboard* and *Columbia Games* corpora, and Stoyanchev & Stent, 2009's from *Let's Go*, a real-world bus information system). In laboratory studies of referential communication and lexical choice, we and others have demonstrated effects of psycholinguistic processing in interactive dialogue contexts,

attempting to gain sufficient control to test hypotheses while modeling spontaneous interaction in conversational contexts. However, the question sometimes arises as to whether dialogue in non-laboratory contexts may be too noisy to show the same sorts of effects (Stoyanchev & Stent, 2009).

In this project, we examine lexical entrainment and conceptual pacts both outside of and inside the laboratory.

The Walking Around Project

Design and Materials

Visual Context Manipulation Visual evidence relevant to a collaborative task has been shown to affect task strategies, to incrementally shape the form and content of spontaneous utterances, and to enable partners to shift the initiative or distribute the responsibility for grounding utterances (see Brennan, 2004 for discussion).

The amount of visual context provided to the stationary direction-giving partner during the navigational task was varied as a between-pairs and within-destinations factor. For half of the target destinations, the direction-giver saw only a close-up photo of the target (Figure 1) along with its numbered location on a map; for the other half, they saw the close-ups plus extra visual context (Figure 2). Each pair was assigned to either List 1 or List 2, with Givers seeing minimally illustrated targets in List 1 or extra visual context in List 2, and vice versa.

Because the mobile partner in the navigation task had rich visual information from being present within the local context, while the stationary partner had global information about the target location but less information about its visual appearance and relative local landmarks, we predicted that additional visual context for the stationary partner would help them align their knowledge, making presentation and acceptance phases of the directions more efficient jointly. On the other hand, such knowledge might be unnecessary for direction-givers who knew campus well (assessed during the final questionnaire).

Methods

Task Sequence Pairs of volunteers did a sequence of five tasks during a 2-hour session: (1) a collaborative navigation task in which one partner followed the other's directions in order to visit 18 pre-chosen destinations on campus, (2) a timed mental rotation test of spatial ability, (3) individual recall tests of their memory for the locations they had discussed, (4) showing the Experimenter the photos taken of the destinations (done by the mobile partner only) during which the Experimenter provided competing labels for some of the destinations, (5) 6 rounds of a referential communication task in which pairs matched identical copies of this set of photos (switching director-matcher roles halfway through, with the mobile partner acting as director for the first three rounds and the stationary partner as director for the last three), and (6) a questionnaire, completed individually. This sequence of tasks enabled us

to examine lexical choices within and across different interactive and solitary contexts; the spatial ability test provided the opportunity to examine the impact of this individual difference on strategies and performance.

Status of Corpus Project The corpus being released to the research community includes 36 digitally recorded spoken dialogues (with associated data on individual differences and experimental parameters). The navigational dialogues from these pairs have all been transcribed in detail (including disfluencies and pauses, along with time stamps). A full article in preparation will present data on the content and sequencing of referring expressions and navigation strategies. Here, we report findings from the post-navigation recall test, entrainment data from the 6-round referential communication task, and effects of spatial ability.

Figure 1: Close-up views of 10 targets seen by Givers *(L to R, top to bottom)*: Patriotic faces, Mushroom house, Outside chalkboard, Ship sculpture, Sorority rock with girl, Spaceship label, Goldfarb plaque, Cylindrical cement structure, Cedar plaque, Warning sticker

Subjects

54 pairs of Stony Brook students were recruited from the Psychology Subject Pool or via flyers on campus to participate with another person in a task involving a *nature hike* or *walking around campus*. All identified as native speakers of English, all but two pairs were strangers, and all gave informed consent. As they arrived, they were asked whether they preferred to walk around or remain indoors; roles were assigned according to their preferences.

Because the sessions involved so many tasks (with the inevitable risks of equipment failure, bad weather, and subjects who either failed to complete the navigation task or who disregarded instructions and ran errands during the

session), and since inclusion of the navigation dialogues in the final corpus required both members of a pair to sign an additional release form, 18 of the 54 sessions could not be included in the Walking Around corpus. However, a total of 49 pairs completed most or all of the matching rounds, so contributed usable data for the current analyses. Subjects received their choice of either 2 research credits that could be used to fulfill a course requirement, or $9 per hour.

The student assigned to the role of direction-Giver (G) remained inside the lab with a landline phone to direct the student assigned as Follower (F) to locations on campus.

Figure 2: Extra visual context provided to Giver (without highlight) for selected targets from Fig. 1: Spaceship label, Warning sticker, Sorority rock with girl, Patriotic faces.

Procedure

The sequence of tasks unfolded as follows:

(1) The goal of the collaborative navigation task was for F to visit and identify each target destination in order, and to take a photograph documenting the destination. F was provided with a digital camera and a mobile phone with GPS, but no information about the destinations, and so had to rely entirely on G's directions and descriptions. G used an interactive web interface displaying a custom Google Map of the relevant portion of Stony Brook's campus, with 18 target destinations marked and numbered in the order in which they were to be visited by F. By clicking on each numbered destination on the map, G could view the associated pop-up photograph(s) and label (Table 1 lists the labels provided to G). In addition, the same photographs and labels were provided as hard copy in a binder, with the materials for one destination visible per page. The map interface also displayed a cursor generated by F's mobile phone's GPS tracking, providing G with visual evidence of F's location on campus, updated every few seconds.

(2) Immediately following the navigation task, F returned to the lab, where both F and G were tested individually for their spatial ability via a timed paper and pencil mental rotation test (*Card Rotation Test–S-1 [Rev.]* by Educational

Testing Services, 1962, 1975).

(3) Then each partner was tested individually for their memory for the destinations. A SuperLab program on a laptop randomly selected and displayed 36 pictures of campus destinations, one at a time; half of these were the close-ups of the eighteen target destinations from the navigational task (as in Figure 1), and the other half depicted destinations the partners had not seen. Each partner was instructed to indicate within five seconds whether the destination was old (discussed or visited in the navigation task) or new. If a partner recognized the destination as old, the program allowed 30 seconds for typing its name into an expanding textbox. This task aimed to probe for the most available conceptualization (presumably, related to the one they had grounded with the partner during the navigational study completed only a few minutes earlier).

(4) Next, the Experimenter attempted to interfere with any conceptual precedents or pacts the pair may have established by giving F printouts of the photos F had just taken and having F identify them one by one; for some, the Experimenter deliberately used different labels than had been originally provided to G (see Table 1).

Table 1: Influences upon G and F, prior to matching task

Label provided to Giver (screen and binder):	Expression addressed to Follower (by Exptr)
Cedar plaque	*Lebanon plaque*
Goldfarb plaque	*Stickball plaque*
Patriotic faces	*Profiles plaque*
Warning sticker	*Watertower sticker*
Mushroom house	*Brown brick structure*
Outside chalkboard	*Physics chalkboards*
Ship sculpture	*Engineering sculpture*
Spaceship label	*SUNYLab label*
Cylindrical cement structure	*T-shaped cement block*
Sorority rock with girl	*Sorority rock with the moon*

(5) After this, the partners re-joined one another to participate in 6 rounds of a referential communication task in which they repeatedly matched identical sets of pictures of the 18 destinations. The pictures used in this phase consisted of the actual photos taken by F while walking around and were printed in grey-scale, on 8.5" X 11" paper. G and F sat at separate tables, back-to-back; F played the role of director for the first 3 rounds with G as matcher, and they switched roles for the last 3 rounds. To begin each round, each partner's set of pictures was shuffled randomly, with the director's arranged as a stack from which to view and describe one picture at a time, and the matcher's spread out on the table so that the matcher could identify and stack each one in the same order as the director's. Partners were told to accurately order their pictures as quickly as possible; the experimenter intervened if they made a matching error or tried to peek at each other's pictures.

(6) Partners filled out a questionnaire asking about their familiarity with campus, whether they knew each other, their length of time at SBU, whether they drive or live on

campus, their confidence in their own sense of direction and ability to read maps, and any other information they wanted to volunteer as relevant. Then they were debriefed.

Coding and Examples

Entrainment coding We coded the content of referring expressions that both partners used in the 6 rounds of the matching task, including whether F's and G's expressions had converged by Round 6 (F was director and G was matcher for Rounds 1-3, with G as director and F as matcher for rounds 4-6). We coded only the 10 target destinations for which the Experimenter provided an expression to F before the matching task, where the expression was distinct enough to compete with G's given label (illustrated in Figure 1 and listed in Table 1).

Lexical convergence (as evidence for entrainment) was coded for each of the 10 pictures using moderately strict criteria (adapted from Brennan & Clark, 1996 and Bortfeld & Brennan, 1997), as follows: if all of the content morphemes (nouns, adjectives) in the expressions spoken by the director (G) in Round 6 were included in any expressions from Rounds 1-3 (where F was director), *or* were also spoken or confirmed (e.g., by a backchannel or verbal acknowledgment) by F (the matcher) in Rounds 4-5, that counted as convergence.

Examples By these criteria, the expression "mushroom house" (used by a particular G as director in Rounds 4-6) was coded as *not* convergent with "mushroom hut" (which the F partner had used throughout Rounds 1-3) because F never used "house" and G never used "hut").

It has been shown many times that entrained-upon expressions become more efficient with repeated referring, so by this criteria, "mushroom" produced during Round 6 was coded as a convergent, shortened form of "brick mushroom" from Rounds 1-3.

Our coding ignored the order of propositional content, as well as the appearance of proxy terms (e.g., "thingy" in "mushroom thingy" or "blah blah" in "environmental blah blah" were considered to be like a wild card or variable, replaceable with other terms). Note that we did not include prepositions or other function words in the coding, so "moon rock girl" was coded as convergent with "girl on the rock with the moon."

However, when G introduced a new propositional morpheme in Round 6 that F had not used or acknowledged in Rounds 1-5, we coded this as a distinct expression (and a failure to entrain), such as in the following example, for which the given label was *Ship sculpture*, and the Experimenter's term was *Engineering sculpture*. (Note: in this and other examples, matchers did not always verbally respond, but would often just identify and arrange a picture.)

Round 1 F: boat structure
       G: boat structure?
       F: yeah, the- in front of the Engineering building

Round 2 F: structure in front of Engineering building, the boat
Round 3 F: structure in front of Engineering building
Round 4 G: wire ship, *the*
       F: *what?*
       G: in front of the-
       F: Engineering building?
       G: yeah
       F: okay
Round 5 G: the ship thing, in front of Engineering
Round 6 G: wire ship   *(Pair 13, #13)*

In the example above, F's "boat structure" was never taken up by G, and G's "wire" was never confirmed by F.

Note that when pairs failed to entrain on a referring expression by Round 6, this did not mean that the expressions they used were unrelated. In some cases, they had all but converged in the 5th round, but introduced an element new to their conversation in the 6th round (which could not count as entrainment by our strict criteria). This is the case with the next example (note also that in the navigation task, G had told F that he hated this particular object because he had run into it on his bicycle as a child):

Round 1: F: the cylinder you ran into
Round 2: F: the cylinder you ran into
Round 3: F: cylinder you ran into
Round 4: G: cylinder
       F: cylinder
Round 5: G: cylinder
Round 6: G: cement   *(Pair 54, #15)*

Note that for this destination, both the label given in the binder and the Experimenter's term had included "cement" (see Table 1). Like other behaviorally expressed products of human memory, conceptual pacts are subject to interference.

In other cases where pairs did not entrain, they sometimes appeared to be in the process of converging on a conceptual pact that was still a bit unstable, having included all of the same elements but in partial combinations during the 6 rounds, with F never producing or accepting all of the propositions in G's final expression.

However, at other times, F and G did express distinct, competing conceptualizations (possibly expressing disagreement), such as when F repeatedly used "submarine" in Rounds 1-3 and G subsequently used "spaceship" on Rounds 4-6, with neither comment nor discussion. Such dueling perspectives rarely led to errors in the matching task, but they do illustrate that pairs did not always succeed in entraining on a stable perspective.

Some gave explicit evidence that they were monitoring the sources of the content they expressed, both from walking around and from hearing competing terms from the Experimenter. In this example (coded as entrainment), F refers explicitly to both:

Round 1: F: the thing I couldn't find, that really tall, I think

she <*the Experimenter*> described it as a
watertower, the really tall thing that was by –
G: - the helium (tank)?
F: yeah, yeah, yeah
Round 2: F: the helium tank, I think you said
Round 3: F: helium, tank, or
Round 4: G: the helium tank
Round 5: G: helium tank
F: ( ) tank
Round 6: G: helium tank   *(Pair 30, #6)*

Other Influences on Perspective Taking   We also coded the extent to which the expressions in Round 6 included content that was (a) originally provided in the label provided to G in the binder for the Walking Around task, (b) spoken to F by the Experimenter after the recall test and before the matching task, or (c) "other", or not provided in (a) or (b). We predict that "other" content of the referring expressions will reflect conceptualizations and expressions that the pair discussed during the navigation task, as here:

Round 1: F: the T-shaped structure that you thought
            looked like a thumbtack
Round 2: F: T-shaped structure, thumbtack
Round 3: F: the thumbtack scri-, structure
Round 4: F: thumbtack
Round 5: F: thumbtack
Round 6: F: thumbtack   *(Pair 49, #15)*

Because we have not yet coded all of the destination-relevant expressions produced in over 30 hours of recorded navigation sessions, we cannot yet quantify the influence of that task on the expressions pairs used later on. So for the purpose of the current analyses, we use these proxies:

- Often, G began describing a destination in the Walking Around task by first proposing the label given for that target in the destination binder. So for the current analyses, we used the initial labels from G's binder as an estimate of the conceptualization first proposed by G and often elaborated, modified, or renegotiated and replaced as F walked around.
- Likewise, we considered content from alternative expressions directed at F by the Experimenter as potentially interfering with conceptual pacts established while walking around (especially due to the recency of that label, heard immediately before the matching task).
- Finally, content coded as "other" was hypothesized as having as its potential source the perspectives negotiated by the partners while walking around.

Results and Discussion

Effects of Spatial Ability   The time it took a given pair to complete the Walking Around task was correlated with G's mental rotation ability, $r = -.42$, $p = .004$. Walking Around time was *not* reliably correlated with F's mental rotation ability, $r = -.24$, $p = .111$.

Entrainment while Matching   By Round 6, there was evidence of lexical entrainment 74.8% of the time, with individual pairs entraining on perspectives for 33% to 100% of destinations, and particular target destinations showing entrainment by 54% - 95% of the pairs.

This represents a rather strong degree of entrainment considering that conceptual pacts are theorized to be flexible across contexts, and the multiple contexts in our study differed substantially from one another (e.g., during the matching task, G viewed F's photos for the first time, and now needed to compare and distinguish the destinations alongside one another. Brennan & Clark's (1996) lab experiments showed 56%-71% entrainment (this did not include expressions where one was a shortened version of the other, but the task context was far more uniform).

Hybrids   When several perspectives are in the air, one highly effective strategy is to hybridize them. In the next example, G first read *warning sticker* in the binder. While walking around, F proposed *a big water tank or something*, while G proposed *diamond shaped sticker*. Later, F heard *watertower* from the Experimenter. In the matching rounds, each accommodated the other:

Round 1: F: the watertower with the triangle, diamond I
            mean
Round 2: F: watertower, with the diamond, the colors
Round 3: F: the watertower with the diamond
Round 4: G: the, the watertower sign, triang-, diamond,
            with the colors
Round 5: G: the watertower sign, warning sign
Round 6: G: the diamond watertower sign   *(Pair 49, #6)*

Memory for Destinations & Labels   G was more accurate than F in correctly recalling the destinations discussed in the navigation task (responding "yes" to old items), 90.6% to 80.4%, $t(47) = 3.69$, $p = .001$. G was also marginally faster to respond than F, $t(47) = 1.72$, $p < .10$. This makes sense, as the contexts in which G encoded the destinations were consistent with the context of recall (memory prompts were the same closeup photos from the binder, and G had ample time to look at them as F walked around).

Convergence of Recalled Terms with Given Label   After responding "yes" in the recall test, the G partner produced an expression with the identical content as the F partner 41.7% of the time. The fact that this was so much lower than their entrainment rate in the matching rounds highlights the difference between individual memories and collaboratively-achieved perspectives.

The expression G recalled had all the same propositional content as the given (binder) label 37.5% of the time, and a somewhat shorter label 20.4% of the time. So 57.9% of the time, the label G recalled matched the one G started out with (we do not call this convergence "entrainment" since it does not emerge from interpersonal interaction). G's recalled label *added* propositional content to the given label

only 2.2% of the time (even though they often created hybrid labels later when interacting with F in the matching rounds). This is striking, as it suggests that when they encoded the given label as precedent, they did so very strongly. The rest of G's recalled expressions (30.9%) included no content from the given label.
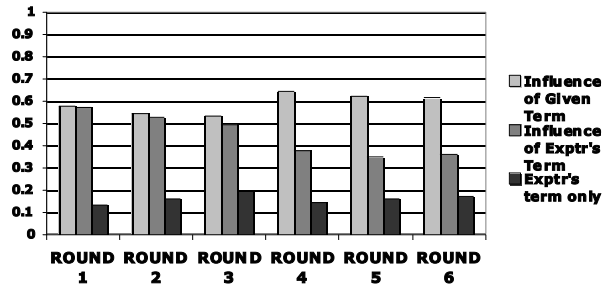

Figure 3: Proportions of matching rounds that included content from labels initially given to G (in G's binder) or spoken to F (by the Experimenter). The discontinuities between Rounds 3 and 4 reflect a change in speaker roles (F directed Rounds 1-3, and G directed Rounds 4-6).

Influence of Experimenter's Label Finally, we consider whether F designed referring expressions with G in mind by quantifying the interference from expressions addressed by the Experimenter to F just before the matching rounds. Figure 3 shows that when F was director (R1-R3), some content from those expressions made it into the expressions F proposed to G (note: this does not yet take into account a baseline for how often such terms were used while walking around). If F chose referring expressions egocentrically or based on recency alone, rather than designing them for G, then the black bars in Figure 3 would be much higher.

Conclusions

The Walking Around Corpus and associated data from this project demonstrate how spontaneous referring expressions in conversation are shaped by partner-specific influences and interaction, as well as by other factors that include prior conceptualizations, spatial ability, and recent interference from another speaker. We are analyzing the referring expressions from the corpus to further explore evidence for entrainment in real world contexts, as well as to address applied questions. The corpus may be applicable to spoken dialogue applications; although GPS-based spoken dialogue systems for cars are now standard, such systems have not yet been optimized for pedestrians. This corpus is available to others for research purposes (contact the first author).

Acknowledgments

References

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory & Language, 46*, 391-418.

Bortfeld, H. & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*, 119-147.

Brennan, S. E. (2004). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action tradition*. Cambridge MA: MIT Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 6*, 1482-1493.

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

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 181-218.

Gergle, D., Kraut, R.E., & Fussell, S.R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction, 28*, 1-39.

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., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1-35.

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

Kosslyn, S. M. (1988). Imagery in learning. In M. S. Gazzaniga (Ed.), *Perspectives in Memory Research*. Cambridge, MA: MIT Press.

Matthews, D., Lieven, E., Tomasello, M. (2010) What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Devel. Psychology, 46*, 749-760.

Metzing, C. & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory & Language, 49*, 201-213.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proc. of the ACL (Short Papers).* 169-172.

Schober, M. F., & Carstensen, L. L. (2010). Does being together for years help comprehension? In E. Morsella (Ed.), *Expressing oneself/Expressing one's self: Communication, cognition, language, and identity*. London: Taylor & Francis.

Stoyanchev, S. & Stent. A. (2009). Lexical and syntactic adaptation and their impact in deployed spoken dialog systems. In *Proc. of NAACL HLT* (Short Papers), 189-192.