Computational Approaches to the Production of Referring Expressions: Dialog Changes (Almost) Everything

Amanda J. Stent
AT&T Labs - Research
Florham Park, NJ 07932, USA
stent@research.att.com

Introduction

The referring expression production task is currently a growing area of research in computational linguistics. Many computational approaches to this task are motivated by psycholinguistic research. However, most of these approaches model only the speaker, not the hearer or the collaborative nature of reference in dialog. In this talk we present an argument for computational research on the production of referring expressions in dialog that is motivated by psycholinguistic findings. First, we survey computational research on the production of referring expressions. Then we present several dialog-related phenomena that affect referring expression production. We sketch a data-driven computational model for coreference in dialog that covers comprehension and production of referring expressions and can be used to examine these dialog-related phenomena. We conclude by presenting some other ideas for research on the production of referring expression in dialog.

Computational Approaches to Production of Referring Expressions

Most computational research on the production of referring expressions looks only at natural language descriptions whose primary function is to *identify* a target referent present in the (linguistic, visual, task or other) discourse context (Reiter & Dale, 1992), as opposed to introducing an entity into the discourse, describing an attribute of an entity (Cheng et al., 2001), or other possible discourse goals. In addition, because most referring expressions are produced by and for humans, most computational research on the production of referring expressions has been evaluated in terms of *humanlikeness*, i.e. similarity of the output to referring expressions produced by humans (Reiter & Dale, 1992; Viethen & Dale, 2006; Gatt, Belz, & Kow, 2008, 2009). Although humanlike referring expressions could be produced in many ways, most researchers motivate their approaches at least in part by appeals to psycholinguistic realism. In fact, the most influential computational model of the production of referring expressions, the incremental algorithm proposed by Dale and Reiter (1992; 1995), is based on a psycholinguistic argument: that the human language production system is subject to memory and processing constraints that may prevent it from producing *optimal referring expressions*[1], causing overgeneration. However, the algorithm is also computationally attractive because it is efficient and domain-independent.

The incremental algorithm works as follows: given a target referent (e.g. a large blue skirt), a set of *distractors* (other possible targets of reference accessible in the discourse context, e.g. large and small shirts and trousers in blue and green and large skirts in blue and green) and an ordered list of *attributes* (e.g. (size, color, type)), the incremental algorithm repeatedly adds an attribute from the list to the description until all distractors are ruled out by the description (here, *the large blue skirt*, where the optimal referring expression is *the blue skirt*). In the incremental algorithm, the attribute list may be ordered once-per-domain (Reiter & Dale, 1992) or once-per-speaker (Bohnet, 2008; Di Fabbrizio, Stent, & Bangalore, 2008; Gatt et al., 2011; Koolen et al., 2009), or may be re-ordered on the fly based on metrics such as priming (Di Fabbrizio et al., 2008; Goudbeek & Krahmer, 2011; Gupta & Stent, 2005), or discriminative power with respect to the distractors (Dale, 1989; Siddharthan & Copestake, 2004). In addition, the distractor set is commonly chosen based on linguistic context (Reiter & Dale, 1992), but may be chosen based on visual context (Guhe, 2009; Gatt et al., 2008, 2009). Originally created only for references to single targets, the incremental algorithm has been extended to cover generation of references to sets of entities (Gatt & Deemter, 2007).

The incremental algorithm requires a rich representation of the domain and context of referring expression production (Reiter & Dale, 1992). This is not generally a realistic requirement; Siddharthan and Copestake have presented a modified version that does not require a domain model (2004).

As Belz and Varges point out in (2007), the incremental algorithm is intended for the production of definite NPs referring to discourse-old referents. However, it has also been applied to the task of generating referring expressions to discourse-new referents (Krahmer & Theune, 2002; Gatt et al., 2008, 2009). Other researchers have addressed the production of pronominal referring expressions and reduced (non-distinguishing) definite NPs, typically using algorithms based on models of discourse salience and focus, e.g. the centering model of local discourse structure (Dale, 1990; Krahmer & Theune, 2002; Passonneau, 1996), or simple heuristics such as recency, syntactic subjecthood and lack of recent competitors (Callaway & Lester, 2002; Greenbacker & McCoy, 2009; Henschel, Cheng, & Poesio, 2000; McCoy & Strube, 1999; Khudyakova et al., 2011). These algorithms augment rather the replace the incremental algorithm.

[1] Although the algorithm is motivated through a psycholinguistic argument, *optimal referring expression* is defined in computational terms, as the shortest referring expression that uniquely identifies the target referent.

Still other research has addressed questions pertaining to the realization of referring expressions, (Bohnet, 2008; Di Fabbrizio et al., 2008; Gupta & Stent, 2005; Krahmer et al., 2008), including situations where the realization of an attribute for an entity, as well as its selection, may depend on the set of distractors (van Deemter, 2000, 2006).

In the last five years the generation community has begun to develop shared tasks with shared resources; this has facilitated both empirical evaluation and data-driven approaches. The first four of these shared tasks all pertained to referring expression generation: ASGRE (Attribute Selection for Generating Referring Expressions, the question addressed in (Dale & Reiter, 1995)) and its successor shared task TUNA (which added surface realization) (Gatt et al., 2008, 2009); GREC (Generating Referring Expressions in Context) (Belz, Kow, & Viethen, 2009; Belz & Kow, 2010; Gatt & Belz, 2008); and GIVE (Giving Instructions in Virtual Environments) (Byron et al., 2009; Koller et al., 2010). This has caused a rapid increase in computational research on the production of referring expressions: the number of papers mentioning 'referring expression generation' in the ACL Anthology[2] is 3 prior to 1992 (the year in which the incremental algorithm was first published), 58 from 1992-2007 (the year of the ASGRE shared task), and 73 from 2008-2011. Twelve papers in the anthology mention ASGRE, 34 mention TUNA, 29 mention GREC, and 18 mention the GIVE challenge.

Dialog Changes (Almost) Everything

In the previous section, we briefly surveyed computational research on referring expression generation. With few exceptions, this research has focussed on the production of referring expressions in non-interactive contexts. One consequence has been a focus on speaker-related constraints: algorithms are evaluated on the basis of how well they match a human speaker's choices (Gatt et al., 2008, 2009; Belz et al., 2009; Belz & Kow, 2010), and most variations on the standard algorithm have been to reflect speaker-related factors, such as speaker-specific preferences and priming (Bohnet, 2008; Di Fabbrizio et al., 2008; Koolen et al., 2009).

Dialog forces us to think more globally about the process of reference. In dialog, the influence of the hearer of a referring expression is inescapable; evaluation must take into account the ease with which a hearer can successfully resolve referring expressions, and algorithms for referring expression production must take into account hearer-related constraints. In this section, we briefly survey important aspects of referring expression production in dialog.

Reference is **collaborative**. A realized description is in fact an *attempt* at a referring action; the actual referring action cannot be completed unless the hearer identifies the intended referent (Heeman & Hirst, 1995). This is true of all reference, but in dialog the evidence of failure of reference attempts is typically both immediate and obvious. Models of reference in dialog must therefore tightly integrate the production and comprehension processes; it is typically not possible to have one without the other (Grosz, 1978; Heeman & Hirst, 1995).

Reference attempts are **tentative**. A corollary of the collaborative nature of reference is that, from an individual speaker's perspective, every reference attempt is tentative until confirmed by the hearer. In other words, grounding is key to successful reference (Clark & Brennan, 1991). Sometimes grounding is explicit, as in this example[3]:

Example 1 (explicit grounding):
A: catalog number of the first item please
B: it is just one thing, a 7 2 0 dash 2 9 1 3
A: 3 that began with a as in alice

Explicit grounding can be verbal or non-verbal; for example, in the GIVE challenge only the instruction giver can speak, but the instruction follower can provide evidence of success of a reference attempt by moving in the world, or evidence of failure by hesitating. However, often grounding is implicit – the hearer only provides evidence of the failure of a reference attempt, and otherwise the speaker can assume that the attempt was at least partially successful, i.e. that the hearer was able to identify a set of potential referents with a degree of confidence and specificity sufficient to enable continued interaction. Compare:

Example 2 (implicit grounding):
B: [uh] i would like that in the antique ivory
A: and how many?

Example 3 (evidence of misunderstanding):
A: that is a brief tall package of two size 40 white totals one
A: next item?
B: could you repeat that was what tall?

Production and comprehension of grounding behavior in dialog is only possible if computational models can track the participants' different understandings of common ground as the dialog progresses.

Reference is **incremental**. Because referring attempts are tentative and (at least in human-human conversation) feedback is fairly immediate, the process of constructing a referring expression should ideally be incremental (DeVault et al., 2005). The incremental nature of reference is most evident when a hearer completes a referring expression started by a speaker, as in the following examples:

Example 4 (completion):
A: what was that number, t a 0 0 7
B: 0 0 4

Example 5 (completion with correction):
A: are you still at [uh]
B: sigfreid
A: sigfield drive

[2]http://aclweb.org/anthology-new/

[3]All the examples in this section are taken from the dialog corpus presented in (Bangalore & Stent, 2009; Stent & Bangalore, 2010).

This type of behavior is less likely to occur in human-computer dialog, but some emerging computational models of dialog permit incremental processing, e.g. (Schlangen & Skantze, 2011; DeVault, Sagae, & Traum, 2009).

Reference is **multi-functional**. Reference in dialog is collaborative behavior in service of larger communicative goals. Consequently, although a referring expression must be primarily interpreted as an attempt to identify the intended referent, a speaker may communicate information about other goals through reference. For example, a speaker may produce a referring expression with the dialog-related purpose of *informing* the hearer (introducing the referent into the discourse context), and the hearer may then produce a different referring expression in order to *acknowledge, clarify* or *query* the first one (Brennan, 2000; Jordan & Walker, 2005):

Example 6 (clarification):
B: i did not receive a catalog which i always usually get
A: which catalog is that madam?
*B: *the latest one* i guess*
*A: *the fall and winter catalog*?*
B: yeah

Speakers may also produce referring expressions for task-related goals, such as to highlight a contrast between two entities, to motivate the hearer to take an action with respect to the intended referent, to summarize, or to express an opinion about the intended referent (Jordan & Walker, 2005):

Example 7 (summarization):
A: and the next item?
B: [uh] c n 8 4 3 6 1 2 9 walking length size 16 white
A: that is a belted short size 16 in the white

Example 8 (expression of opinion):
A: would you be interested in a [um] paul sebastian cologne spray?
A: would be a good gift for father's day...

As these examples show, the extra-referential intentions of the speaker may strongly influence both the choice to refer and the form of referring expressions. Computational models for referring expression production should therefore take into account information from intentional models of dialog processing (e.g. (Bangalore & Stent, 2009)).

Finally, reference is **hearer-oriented**. The choice to make a reference attempt, the extra-referential intentions of the speaker, and other speaker-related features will have a large impact on the form of a referring expression. However, because reference is collaborative no reference attempt can succeed without the understanding and agreement of the hearer. Consequently, the decisions of speakers to make reference attempts, and their choices about the content of those attempts, will be to at least some degree hearer-oriented (Heller, Skovbroten, & Tanenhaus, 2009; On Yoon, Koh, & Brown-Schmidt, 2011). For example, in a navigation task a speaker may make a reference to a landmark if a hearer seems confused. In an instruction giving task a speaker may make a referring expression less technical if the hearer is not a domain expert (Janarthanam & Lemon, 2010). In a map task, speakers and hearers may converge on a particular perspective (Stoyanchev & Stent, 2007).

In addition to hearer-oriented choices in the content of referring expressions, speakers may align the form of referring expressions to the hearer's previously expressed semantic, syntactic and lexical choices (Brennan & Clark, 1996; Goudbeek & Krahmer, 2010, 2011; Rij, Rijn, & Hendriks, 2011). This reinforces the need for computational models of referring expression comprehension and production to be tightly integrated at all levels of linguistic processing (Buschmeier, Bergmann, & Kopp, 2010; Stent, 2011).

The hearer-oriented nature of reference in dialog also has implications for the evaluation of models of referring expression production. In addition to measuring how similar automatically generated referring expressions are to those produced by humans, researchers can and should evaluate how well referring expressions facilitate comprehension. For example, one could measure how often automatically generated referring expressions can successfully be resolved, how quickly they can be resolved, and whether they cause the hearer to present evidence of confusion (e.g. hesitating, asking clarification questions) (Byron et al., 2009; Gatt et al., 2009; Koller et al., 2010).

The factors influencing reference outlined here would seem to impose heavy knowledge representation and reasoning requirements on computational models of reference in dialog, including explicit tracking of grounding status, speaker's dialog- and task-related intentions, and lexical and syntactic elements of the discourse context (for alignment). However, computational models of dialog already successfully infer and track approximations of many of these elements; for example, dialog acts are approximate representations of speaker intentions, and adjacency pairs are approximate representations of grounding behavior (e.g. (Bangalore & Stent, 2009)). Consequently, it is possible to build models of reference for dialog that take into account some of these factors.

A Joint Model for Comprehension and Production of Referring Expressions in Dialog

In this section we present a model for reference in dialog that permits us to examine some of the aspects of reference outlined above. We are engaged in a long-term research project aimed at designing methods for training models of dialog from human-human data that can be used: (a) to test theories of dialog structure and its impact on production and comprehension of dialog contributions; (b) directly, as components of dialog systems (Bangalore, Fabbrizio, & Stent, 2008; Bangalore & Stent, 2009). The model for reference we propose here is part of this long-term effort. In our talk, we will present worked examples using this model, and discuss our current work in this area.

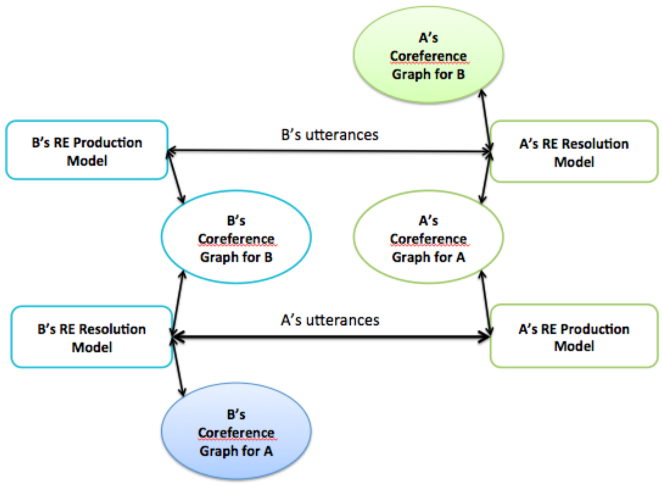

Figure 1: A joint model for comprehension and production of referring expressions in dialog

Publicly available corpora of reference in dialog are small (Anderson et al., 1991; Jordan & Walker, 2005) or not task-oriented (Calhoun et al., 2010; Luo, Florian, & Ward, 2009; Manually Annotated Sub-Corpus, 2011). This has limited the ability of researchers to train computational models of reference in dialog from data or to test the full range of factors that influence referring expression production (Jordan & Walker, 2005; Gupta & Stent, 2005). As a result of a large annotation effort, we have access to a corpus of over 800 human-human task-oriented dialogs in a catalog ordering domain which have been transcribed and labeled with dialog acts, task/subtask information, and, recently, coreference information (Bangalore et al., 2008; Stent & Bangalore, 2010). We use this corpus in experiments with our model.

As depicted in Figure 1, our model includes three speaker-specific components for each speaker: a referring expression resolution component, a referring expression production component, and a coreference graph.

Referring expression resolution The referring expression resolution component for each speaker is a cluster-ranking model (Rahman & Ng, 2009), with additional dialog- and task-specific features (Stent & Bangalore, 2010). In cases where the dialog model has access to a model of the domain, the referring expression resolution model can alternatively be an entity-mention model (Rahman & Ng, 2009).

Referring expression production The referring expression production component for each speaker has two parts. The first part, the referring expression type predictor, uses an algorithm similar to that in (McCoy & Strube, 1999) to determine the type of each output referring expression (pronoun, deictic, definite NP, indefinite NP, proper noun). The second part, the referring expression realizer, selects attributes for definite and indefinite NPs and performs surface realization using a method similar to that in (Siddharthan & Copestake, 2004).

Coreference graph The coreference graph for each speaker contains that speaker's understanding of the coreference links in the discourse so far. In the most basic form of modeling, the graph is updated and used without regard to grounding, priming, or alignment. When a hearer attempts to resolve a referring expression RE_j, links between RE_j and referring expression clusters possibly coreferent with RE_j are added to its coreference graph; these links are weighted according to their rank in the output of the referring expression resolution component. When a speaker produces a referring expression RE_j designed to be coreferent with a referring expression cluster c_{RE_j}, a link of weight 1 connecting RE_j and c_{RE_j} is added to its coreference graph.

Each dialog participant may also keep a coreference graph for the dialog partner; in this case, links are added based on the participant's interpretation model for each referring expression produced by either dialog participant. This coreference graph can be used during referring expression production to select from multiple possible referring expressions the one most likely to be successfully resolved.

Modeling grounding, priming and alignment The coreference graph thus enables us to track the tentative nature of reference, as well as hearer-oriented influences on referring expression production.

The weights of links in the coreference graph can be updated using a model of grounding. When a link between a referring expression RE_j and coreference cluster c_{RE_j} is explicitly grounded, its weight can be increased and the weight of links from RE_j to other coreference clusters can be decreased. On the other hand, when a link is explicitly rejected, it can be removed from the coreference graph and its weight reassigned. This permits a hearer to avoid committing to a particular referent for a referring expression until it is required to; for example, when it has to refer to the cluster itself.

The coreference graph can also be used to track priming and alignment: for example, a speaker can use highly weighted links to recent referring expressions during attribute selection and surface realization, as in (Gupta & Stent, 2005).

Modeling multi-functional reference We can examine the impact of modeling the multi-functional nature of reference by varying how we train the speaker-specific resolution and production components.

When we train the referring expression resolution and production components of the model, we use dialog- and task-related features including dialog act, task/subtask, and intentional state (Bangalore & Stent, 2009; Stent & Bangalore, 2010). This allows the model to track how reference varies by communicative intent.

We can also vary how we train the models to explore other speaker- and hearer-specific aspects of reference. For example, task-oriented dialog, the partners often play different roles, e.g. instruction giver/instruction follower, tutor/student, agent/customer. To test theories about how reference may vary based on speaker role, we can train the resolution and production components using all our training data, or only the training data for the speaker role being targeted.

Using the model This model can be used in a simulation setup such as the one in (Bangalore & Stent, 2009), to test alternatives for representing and reasoning about reference. The best-performing components for the participant to be played by the system can then be used in a dialog system.

This model is incremental only at the level of individual utterances. This means that we cannot use it to look at collaborative construction of individual referring expressions. However, it otherwise permits exploration of all the dialog-related aspects of reference discussed in the previous section.

Discussion

In this talk we have presented an argument for greatly expanding computational research on referring expression in dialog, using findings from the psycholinguistic literature. We conclude by briefly outlining three areas for computational research on the production of referring expressions in dialog that are somewhat tangential to the model presented here.

Use of low-level processing information Psycholinguists are taking advantage of very fine-grained low-level information about speakers' production processes and hearer's comprehension processes, such as eye gaze (Bard, Hill, Arai, & Foster, 2009; Heller et al., 2009). In some computational research on referring expression production in dialog, e.g. in the GIVE challenge (Byron et al., 2009; Koller et al., 2010), computational linguists have access to similar kinds of low-level timing and processing information. Computational research on incremental referring expression production and comprehension could take advantage of this low-level processing information (Staudte & Crocker, 2009).

Non-collaborative dialog Our research is focussed on collaborative task-oriented dialog. However, dialog partners need not be fully cooperative or fully task-focussed. In particular, a dialog system producing referring expressions may have its own (social or other) agenda, such as to educate a student (Graesser et al., 2008), sell a product (Morik, 1989), or hide information from a third party (Traum, 2008). Each of these situations may impose additional constraints on the generation process. To pick just one example, a dialog system wishing to conceal information from a third-party observer may use highly elliptical referring expressions, designed to be comprehensible to one hearer but not others.

Non-humanlike reference Current research on referring expression production assumes that speakers should produce referring expressions that are humanlike. However, there are dialog applications where humanlikeness may be unnecessary or maladaptive. One broad class would be applications where the intended hearer of the referring expression is not a human, but a computer. For example, an information retrieval system might use automatically generated referring expressions as search and indexing terms for documents (Schäfer et al., 2011). Another broad class would be applications where clarity or brevity in referring expressions are desirable goals; for example, applications that help people with aphasia to construct referring expressions (Nikolova, Tremaine, & Cook, 2010), or applications intended for recipients under high cognitive load, such as medical personnel in a critical care facility (Gatt et al., 2009).

Acknowledgments The work presented in this paper is joint work with several colleagues and former students, including Surabhi Gupta, Svetlana Stoyanchev, Giuseppe Di Fabbrizio and Srinivas Bangalore. We would also like to thank Barbara Hollister and her group for their extensive and careful dialog annotation efforts.

References

Anderson, A., et al. (1991). The HCRC map task corpus. *Language and Speech*, *34*(4), 351–366.

Bangalore, S., Fabbrizio, G. D., & Stent, A. (2008). Learning the structure of task-driven human-human dialog. *IEEE Transactions on Audio, Speech, and Language Processing special issue on New Approaches to Statistical Speech and Text Processing*, *16*(7), 1249–1259.

Bangalore, S., & Stent, A. (2009). Incremental parsing models for dialog task structure. In *Proceedings of EACL.*

Bard, E., Hill, R., Arai, M., & Foster, M. E. (2009). Accessibility and attention in situated dialogue: Roles and regulations. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Belz, A., & Kow, E. (2010). The GREC challenges 2010: Overview and evaluation results. In *Proceedings of INLG.*

Belz, A., Kow, E., & Viethen, J. (2009). The GREC named entity generation challenge 2009: Overview and evaluation results. In *Proceedings of the Workshop on Language Generation and Summarisation.*

Belz, A., & Varges, S. (2007). Generation of repeated references to discourse entities. In *Proceedings of ENLG.*

Bohnet, B. (2008). The fingerprint of human referring expressions and their surface realization with graph transducers. In *Proceedings of INLG.*

Brennan, S. (2000). Processes that shape conversation and their implications for computational linguistics. In *Proceedings of ACL.*

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.

Buschmeier, H., Bergmann, K., & Kopp, S. (2010). An alignment-capable microplanner for natural language generation. In *Proceedings of ENLG.*

Byron, D., et al. (2009). Report on the first NLG challenge on generating instructions in virtual environments (GIVE). In *Proceedings of ENLG.*

Calhoun, S., et al. (2010). The NXT-format Switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, *44*(4), 387–419.

Callaway, C., & Lester, J. (2002). Pronominalization in generated discourse and dialogue. In *Proceedings of ACL.*

Cheng, H., et al. (2001). Corpus-based NP modifier generation. In *Proceedings of NAACL.*

Clark, H., & Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). American Psychological Association.

Dale, R. (1989). Cooking up referring expressions. In *Proceedings of ACL.*

Dale, R. (1990). *Generating Referring Expressions*. Cambridge, MA: MIT Press.

Dale, R., & Reiter, E. (1995). Computational interpretation of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(8), 233–263.

DeVault, D., et al. (2005). An information-state approach to collaborative reference. In *Proceedings of ACL.*

DeVault, D., Sagae, K., & Traum, D. (2009). Can I finish? Learning when to respond to incremental interpretation results in interactive dialogue. In *Proceedings of SIGDIAL.*

Di Fabbrizio, G., Stent, A., & Bangalore, S. (2008). Trainable speaker-based referring expression generation. In *Proceedings of CONLL.*

Gatt, A., & Belz, A. (2008). Attribute selection for referring expression generation: New algorithms and evaluation methods. In *Proceedings of INLG.*

Gatt, A., Belz, A., & Kow, E. (2008). The TUNA challenge 2008: Overview and evaluation results. In *Proceedings of INLG.*

Gatt, A., Belz, A., & Kow, E. (2009). The TUNA-REG challenge 2009: Overview and evaluation results. In *Proceedings of ENLG.*

Gatt, A., & Deemter, K. van. (2007). Incremental generation of plural descriptions: Similarity and partitioning. In *Proceedings of EMNLP-CONLL.*

Gatt, A., et al. (2009). From data to text in the neonatal intensive care unit: Using NLG technology for decision support and information management. *AI Communications*, *22*, 153–186.

Gatt, A., et al. (2011). Non-deterministic attribute selection in reference production. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Goudbeek, M., & Krahmer, E. (2010). Preferences versus adaptation during referring expression generation. In *Proceedings of ACL.*

Goudbeek, M., & Krahmer, E. (2011). Referring under load: Disentangling preference-based and alignment-based selection processes in referring expression generation. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Graesser, A., et al. (2008). AutoTutor: Learning through natural language dialogue that adapts to the cognitive and affective states of the learner. In D. Robinson & G. Schraw (Eds.), *Current Perspectives on Cognition, Learning and Instruction: Recent Innovations in Educational Technology that Facilitate Student Learning.*

Greenbacker, C., & McCoy, K. (2009). Feature selection for reference generation as informed by psycholinguistic research. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Grosz, B. (1978). Focusing in dialog. In *Theoretical Issues in Natural Language Processing 2.*

Guhe, M. (2009). Generating referring expressions with a cognitive model. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Gupta, S., & Stent, A. (2005). Automatic evaluation of referring expression generation using corpora. In *Proceedings of the Workshop on Using Corpora in Natural Language Generation.*

Heeman, P., & Hirst, G. (1995). Collaborating on referring expressions. *Computational Linguistics*, *21*(3), 351–382.

Heller, D., Skovbroten, K., & Tanenhaus, M. (2009). Experimental evidence for speakers sensitivity to common vs. privileged ground in the production of names. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Henschel, R., Cheng, H., & Poesio, M. (2000). Pronominalization revisited. In *Proceedings of COLING.*

Janarthanam, S., & Lemon, O. (2010). Learning to adapt to unknown users: Referring expression generation in spoken dialogue systems. In *Proceedings of ACL.*

Jordan, P., & Walker, M. (2005). Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research*, *24*, 157–194.

Khudyakova, M., et al. (2011). Computational modeling of referential choice: Major and minor referential options. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Koller, A., et al. (2010). Report on the second NLG challenge on generating instructions in virtual environments (GIVE-2). In *Proceedings of INLG.*

Koolen, R., et al. (2009). Need i say more? on factors causing referential overspecification. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Krahmer, E., et al. (2008). GRAPH: The costs of redundancy in referring expressions. In *Proceedings of INLG.*

Krahmer, E., & Theune, M. (2002). Efficient context-sensitive generation of referring expressions. In *Information Sharing: Reference and Presupposition in Language Generation and Interpretation.* Stanford, CA: CSLI.

Luo, X., Florian, R., & Ward, T. (2009). Improving coreference resolution by using conversational metadata. In *Proceedings of NAACL.*

Manually annotated sub-corpus. (2011). Available from `http://www.anc.org/MASC/`

McCoy, K., & Strube, M. (1999). Generating anaphoric expressions: Pronoun or definite description. In *Proceedings of the ACL Workshop on Reference and Discourse Structure.*

Morik, K. (1989). User models and conversational settings: Modeling the user's wants. In A. Kobsa & W. Wahlster (Eds.), *User Models in Dialog Systems.* New York: Springer-Verlag.

Nikolova, S., Tremaine, M., & Cook, P. (2010). Click on bake to get cookies: guiding word-finding with semantic associations. In *Proceedings of ASSETS.*

On Yoon, S., Koh, S., & Brown-Schmidt, S. (2011). Influence of perspective and goals on reference production in conversation. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Passonneau, R. (1996). Using centering to relax Gricean informational constraints on discourse anaphoric noun phrases. *Language and Speech*, *39*(2–3), 229–264.

Rahman, A., & Ng, V. (2009). Supervised models for coreference resolution. In *Proceedings of EMNLP.*

Reiter, E., & Dale, R. (1992). A fast algorithm for the generation of referring expressions. In *Proceedings of COLING.*

Rij, J. van, Rijn, H. van, & Hendriks, P. (2011). Towards a cognitively plausible computational model of reference. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Schäfer, U., et al. (2011). The ACL Anthology searchbench. In *Proceedings of ACL-HLT.*

Schlangen, D., & Skantze, G. (2011). A general, abstract model of incremental dialogue processing. *Dialogue & Discourse*, *2*(1).

Siddharthan, A., & Copestake, A. (2004). Generating referring expressions in open domains. In *Proceedings of ACL.*

Staudte, M., & Crocker, M. (2009). Producing and resolving multimodal referring expressions in human-robot interaction. In *Proceedings of the Pre-CogSci Workshop on Production of Referring Expressions.*

Stent, A. (2011). Shared experiences, shared representations, and the implications for applied natural language processing. In *Proceedings of FLAIRS.*

Stent, A., & Bangalore, S. (2010). Interaction between dialog structure and coreference resolution. In *Proceedings of SLT.*

Stoyanchev, S., & Stent, A. (2007). Measuring adaptation between dialogs. In *Proceedings of SIGDIAL.*

Traum, D. (2008). *Computational models of non-cooperative dialogue.* Available from `http://www.kcl.ac.uk/content/1/c6/03/89/15/papertraum.pdf`

van Deemter, K. (2000). Generating vague descriptions. In *Proceedings of INLG.*

van Deemter, K. (2006). Generating referring expressions that involve gradable properties. *Computational Linguistics*, *32*(2), 195–222.

Viethen, J., & Dale, R. (2006). Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of INLG.*