Predicting Success in Dialogue

David Reitter and Johanna D. Moore
dreitter | jmoore @ inf.ed.ac.uk
School of Informatics
University of Edinburgh
United Kingdom

Abstract

Task-solving in dialogue depends on the linguistic alignment of the interlocutors, which Pickering & Garrod (2004) have suggested to be based on mechanistic repetition effects. In this paper, we seek confirmation of this hypothesis by looking at repetition in corpora, and whether repetition is correlated with task success. We show that the relevant repetition tendency is based on slow adaptation rather than short-term priming and demonstrate that lexical and syntactic repetition is a reliable predictor of task success given the first five minutes of a task-oriented dialogue.

1 Introduction

While humans are remarkably efficient, flexible and reliable communicators, we are far from perfect. Our dialogues differ in how successfully information is conveyed. In task-oriented dialogue, where the interlocutors are communicating to solve a problem, task success is a crucial indicator of the success of the communication.

An automatic measure of task success would be useful for evaluating conversations among humans, e.g., for evaluating agents in a call center. In human-computer dialogues, predicting the task success after just a first few turns of the conversation could avoid disappointment: if the conversation isn't going well, a caller may be passed on to a human operator, or the system may switch dialogue strategies. As a first step, we focus on human-human dialogue, since current spoken dialogue systems do not yet yield long, syntactically complex conversations.

In this paper, we use syntactic and lexical features to predict task success in an environment where we assume no speaker model, no semantic information and no information typical for a human-computer dialogue system, e.g., ASR confidence. The features we use are based on a psychological theory, linking alignment between dialogue participants to low-level syntactic priming. An examination of this priming reveals differences between short-term and long-term effects.

1.1 Repetition supports dialogue

In their *Interactive Alignment Model* (IAM), Pickering and Garrod (2004) suggest that dialogue between humans is greatly aided by *aligning* representations on several linguistic and conceptual levels. This effect is assumed to be driven by a cascade of linguistic priming effects, where interlocutors tend to re-use lexical, syntactic and other linguistic structures after their introduction. Such re-use leads speakers to agree on a common situation model. Several studies have shown that speakers copy their interlocutor's syntax (Branigan et al., 1999). This effect is usually referred to as *structural* (or: *syntactic*) *priming*. These persistence effects are inter-related, as lexical repetition implies preferences for syntactic choices, and syntactic choices lead to preferred semantic interpretations. Without demanding additional cognitive resources, the effects form a causal chain that will benefit the interlocutor's purposes. Or, at the very least, it will be easier for them to repeat linguistic choices than to

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 808–815,
Prague, Czech Republic, June 2007.

actively discuss their terminology and keep track of each other's current knowledge of the situation in order to come to a mutual understanding.

1.2 Structural priming

The repetition effect at the center of this paper, priming, is defined as a tendency to repeat linguistic decisions. Priming has been shown to affect language production and, to a lesser extent, comprehension, at different levels of linguistic analysis. This tendency may show up in various ways, for instance in the case of lexical priming as a shorter response time in lexical decision making tasks, or as a preference for one syntactic construction over an alternative one in syntactic priming (Bock, 1986). In an experimental study (Branigan et al., 1999), subjects were primed by completing either sentence (1a) or (1b):

1a. *The racing driver showed the torn overall...*
1b. *The racing driver showed the helpful mechanic...*

Sentence (1a) was to be completed with a prepositional object ("to the helpful mechanic"), while (1b) required a double object construction ("the torn overall"). Subsequently, subjects were allowed to freely complete a sentence such as the following one, describing a picture they were shown:

2. *The patient showed ...*

Subjects were more likely to complete (2) with a double-object construction when primed with (1b), and with a prepositional object construction when primed with (1a).

In a previous corpus-study, using transcriptions of spontaneous, task-oriented and non-task-oriented dialogue, utterances were annotated with syntactic trees, which we then used to determine the phrase-structure rules that licensed production (and comprehension) of the utterances (Reitter et al., 2006b). For each rule, the time of its occurrence was noted, e.g. we noted

3. 117.9*s NP* $\rightarrow$ *AT AP NN a fenced meadow*
4. 125.5*s NP* $\rightarrow$ *AT AP NN the abandoned cottage*

In this study, we then found that the re-occurrence of a rule (as in 4) was correlated with the temporal distance to the first occurrence (3), e.g., 7.6 seconds. The shorter the distance between prime (3) and target (4), the more likely were rules to re-occur.

In a conversation, priming may lead a speaker to choose a verb over a synonym because their interlocutor has used it a few seconds before. This, in turn, will increase the likelihood of the structural form of the arguments in the governed verbal phrase–simply because lexical items have their preferences for particular syntactic structures, but also because structural priming may be stronger if lexical items are repeated (lexical boost, Pickering and Branigan (1998)). Additionally, the structural priming effects introduced above will make a previously observed or produced syntactic structure more likely to be re-used. This chain reaction leads interlocutors in dialogue to reach a common situation model. Note that the IAM, in which interlocutors automatically and cheaply build a common representation of common knowledge, is at odds with views that afford each dialogue participant an explicit and separate representation of their interlocutor's knowledge.

The connection between linguistic persistence or priming effects and the success of dialogue is crucial for the IAM. The predictions arising from this, however, have eluded testing so far. In our previous study (Reitter et al., 2006b), we found more syntactic priming in the task-oriented dialogues of the Map Task corpus than in the spontaneous conversation collected in the Switchboard corpus. However, we compared priming effects across two datasets, where participants and conversation topics differed greatly. Switchboard contains spontaneous conversation over the telephone, while the task-oriented Map Task corpus was recorded with interlocutors co-present. While the result (more priming in task-oriented dialogue) supported the predictions of IAM, cognitive load effects could not be distinguished from priming. In the current study, we examine structural repetition in task-oriented dialogue only and focus on an extrinsic measure, namely task success.

2 Related Work

Prior work on predicting task success has been done in the context of human-computer spoken dialogue systems. Features such as recognition error rates, natural language understanding confidence and context shifts, confirmations and re-prompts (dialogue management) have been used classify dialogues into *successful* and *problematic* ones (Walker et al., 2000). With these automatically obtainable features, an accuracy of 79% can be achieved given the first two turns of "How may I help you?" dialogues, where callers are supposed to be routed given a short statement from them about what they would like to do. From the whole interaction (very rarely more than five turns), 87% accuracy can be achieved (36% of dialogues had been hand-labeled "problematic"). However, the most predictive features, which related to automatic speech recognition errors, are neither available in the human-human dialogue we are concerned with, nor are they likely to be the cause of communication problems there.

Moreover, failures in the Map Task dialogues are due to the actual goings-on when two interlocutors engage in collaborative problem-solving to jointly reach an understanding. In such dialogues, interlocutors work over a period of about half an hour. To predict their degree of success, we will leverage the phenomenon of *persistence*, or *priming*.

In previous work, two paradigms have seen extensive use to measure repetition and priming effects. *Experimental studies* expose subjects to a particular syntactic construction, either by having them produce the construction by completing a sample sentence, or by having an experimenter or confederate interlocutor use the construction. Then, subjects are asked to describe a picture or continue with a given task, eliciting the target construction or a competing, semantically equivalent alternative. The analysis then shows an effect of the controlled condition on the subject's use of the target construction.

Observational studies use naturalistic data, such as text and dialogue found in corpora. Here, the prime construction is not controlled–but again, a correlation between primes and targets is sought. Specific competing constructions such as active/passive, verbal particle placement or *that*-deletion in English are often the object of study (Szmrecsanyi, 2005; Gries, 2005; Dubey et al., 2005; Jäger, 2006), but the effect can also be generalized to syntactic phrase-structure rules or combinatorial categories (Reitter et al., 2006a).

Church (2000) proposes adaptive language models to account for lexical adaptation. Each document is split into *prime* and *target* halves. Then, for selected words w, the model estimates

$$P(+adapt) = P(w \in target | w \in prime)$$

$P(+adapt)$ is higher than $P_{prior} = P(w \in target)$, which is not surprising, since texts are usually about a limited number of topics.

This method looks at repetition over whole document halves, independently of decay. In this paper, we apply the same technique to syntactic rules, where we expect to estimate syntactic priming effects of the long-term variety.

3 Repetition-based Success Prediction

3.1 The Success Prediction Task

In the following, we define two variants of the task and then describe a model that uses repetition effects to predict success.

Task 1: *Success is estimated* when an entire dialogue is given. All linguistic and non-linguistic information available may be used. This task reflects post-hoc analysis applications, where dialogues need to be evaluated without the actual success measure being available for each dialogue. This covers cases where, e.g., it is unclear whether a call center agent or an automated system actually responded to the call satisfactorily.

Task 2: *Success is predicted* when just the initial 5-minute portion of the dialogue is available. A dialogue system's or a call center agent's strategy may be influenced depending on such a prediction.

3.2 Method

To address the tasks described in the previous Section, we train support vector machines (SVM) to predict the task success score of a dialogue from lexical and syntactic repetition information accumulated up to a specified point in time in the dialogue.

Data

The HCRC Map Task corpus (Anderson et al., 1991) contains 128 dialogues between subjects, who were given two slightly different maps depicting the same (imaginary) landscape. One subject gives directions for a predefined route to another subject, who follows them and draws a route on their map.

The spoken interactions were recorded, transcribed and syntactically annotated with phrase-structure grammar.

The Map Task provides us with a precise measure of success, namely the deviation of the predefined and followed route. Success can be quantified by computing the inverse deviation between subjects' paths. Both subjects in each trial were asked to draw "their" respective route on the map that they were given. The deviation between the respective paths drawn by interlocutors was then determined as the area covered in between the paths (PATHDEV).

Features

Repetition is measured on a lexical and a syntactic level. To do so, we identify all constituents in the utterances as per phrase-structure analysis. *[Go [to [the [[white house] [on [the right]]]]]]* would yield 11 constituents. Each constituent is licensed by a syntactic rule, for instance *VP* $\rightarrow$ *V PP* for the topmost constituent in the above example.

For each constituent, we check whether it is a lexical or syntactic repetition, i.e. if the same words occurred before, or if the licensing rule has occurred before in the same dialogue. If so, we increment counters for lexical and/or syntactic repetitions, and increase a further counter for string repetition by the length of the phrase (in characters). The latter variable accounts for the repetition of long phrases.

We include a data point for each 10-second interval of the dialogue, with features reporting the lexical (LEXREP), syntactic (SYNREP) and character-based (CHARREP) repetitions up to that point in time. A time stamp and the total numbers of constituents and characters are also included (LENGTH). This way, the model may work with repetition proportions rather than the absolute counts.

We train a support vector machine for regression with a radial basis function kernel ($\gamma = 5$), using the features as described above and the PATHDEV score as output.

3.3 Evaluation

We cast the task as a regression problem. To predict a dialogue's score, we apply the SVM to its data points. The mean outcome is the estimated score.

A suitable evaluation measure, the classical r^2, indicates the proportion of the variance in the actual task success score that can be predicted by the model. All results reported here are produced from 10-fold cross-validated 90% training / 10% test splits of the dialogues. No full dialogue was included in both test and training sets.

	Task 1	Task 2
ALL Features	**0.17**	**0.14**
ALL w/o SYNREP	0.15	0.06
ALL w/o LEX/CHARREP	0.09	0.07
LENGTH ONLY	0.09	n/a
Baseline	0.01	0.01

Table 1: Portion of variance explained (r^2)

Task 1 was evaluated with all data, the Task 2 model was trained and tested on data points sampled from the first 5 minutes of the dialogue.

For Task 1 (full dialogues), the results (Table 1) indicate that ALL repetition features together with the LENGTH of the conversation, account for about 17% of the total score variance. The repetition features improve on the performance achieved from dialogue length alone (about 9%).

For the more difficult Task 2, ALL features together achieve 14% of the variance. (Note that LENGTH is not available.) When the syntactic repetition feature is taken out and only lexical (LEXREP) and character repetition (CHARREP) are used, we achieve 6% in explained variance.

The baseline is implemented as a model that always estimates the mean score. It should, theoretically, be close to 0.

3.4 Discussion

Obviously, linguistic information alone will not explain the majority of the task-solving abilities. Apart from subject-related factors, communicative strategies will play a role.

However, linguistic repetition serves as a good predictor of how well interlocutors will complete their joint task. The features used are relatively simple: provided there is some syntactic annotation, rule repetition can easily be detected. Even without syntactic information, lexical repetition already goes a long way.

But what kind of repetition is it that plays a role in task-oriented dialogue? Leaving out features is not an ideal method to quantify their influence–in particular, where features inter-correlate. The contribution of syntactic repetition is still unclear from the present results: it acts as a useful predictor only over the course of the whole dialogues, but not within a 5-minute time span, where the SVM cannot incorporate its informational content.

We will therefore turn to a more detailed analysis of structural repetition, which should help us draw conclusions relating to the psycholinguistics of dialogue.

4 Long term and short term priming

In the following, we will examine syntactic (structural) priming as one of the driving forces behind alignment. We choose syntactic over lexical priming for two reasons. Lexical repetition due to priming is difficult to distinguish from repetition that is due to interlocutors attending to a particular topic of conversation, which, in coherent dialogue, means that topics are clustered. Lexical choice reflects those topics, hence we expect clusters of particular terminology. Secondly: the maps used to collect the dialogues in the Map Task corpus contained landmarks with labels. It is only natural (even if by means to cross-modal priming) that speakers will identify landmarks using the labels and show little variability in lexical choice. We will measure repetition of syntactic rules, whereby word-by-word repetition (topicality effects, parroting) is explicitly excluded.

For syntactic priming[1], two repetition effects have been identified. Classical priming effects are strong: around 10% for syntactic rules (Reitter et al., 2006b). However, they decay quickly (Branigan et al., 1999) and reach a low plateau after a few seconds, which likens to the effect to semantic (similarity) priming. What complicates matters is that there is also a different, long-term adaptation effect that is also commonly called (repetition) priming.

Adaptation has been shown to last longer, from minutes (Bock and Griffin, 2000) to several days. Lexical boost interactions, where the lexical repetition of material within the repeated structure strengthens structural priming, have been observed for short-term priming, but not for long-term priming trials where material intervened between prime and target utterances (Konopka and Bock, 2005). Thus, short- and long-term adaptation effects may well be due to separate cognitive processes, as recently argued by (Ferreira and Bock, 2006). Section 5 deals with decay-based short-term priming, Section 6 with long-term adaptation.

Pickering and Garrod (2004) do not make the type of priming supporting alignment explicit. Should we find differences in the way task success interacts with different kinds of repetition effects, then this would be a good indication about what effect supports IAM. More concretely, we could say whether alignment is due to the automatic, classical *priming* effect, or whether it is based on a long-term effect that is possibly closer to implicit learning (Chang et al., 2006).

5 Short-term priming

In this section, we attempt to detect differences in the strength of short-term priming in successful and less successful dialogues. To do so, we use the measure of priming strength established by Reitter et al. (2006b), which then allows us to test whether priming interacts with task success. Under the assumptions of IAM we would expect successful dialogues to show more priming than unsuccessful ones.

Obviously, difficulties with the task at hand may be due to a range of problems that the subjects may have, linguistic and otherwise. But given that the dialogues contain variable levels of syntactic priming, one would expect that this has at least some influence on the outcome of the task.

5.1 Method: Logistic Regression

We used mixed-effects regression models that predict a binary outcome (repetition) using a number of discrete and continuous factors.[2]

As a first step, our modeling effort tries to establish a priming effect. To do so, we can make use of the fact that the priming effect decays over time. How strong that decay is gives us an indication of how much repetition probability we see shortly after the stimulus (prime) compared to the probability of chance repetition–without ever explicitly calculating such a prior.

Thus we define the strength of priming as the decay rate of repetition probability, from shortly after

[1] in production and comprehension, which we will not distinguish further for space reasons. Our data are (off-line) production data.

[2] We use Generalized Linear Mixed Effects models fitted using GlmmPQL in the MASS R library.

the prime to 15 seconds afterward (predictor: DIST). Thus, we take several samples at varying distances (d), looking at cases of structural repetition, and cases where structure has not been repeated.

In the syntactic context, syntactic rules such as VP $\rightarrow$ VP PP reflect syntactic decisions. Priming of a syntactic construction shows up in the tendency to repeat such rules in different lexical contexts. Thus, we examine whether syntactic rules have been repeated at a distance d. For each syntactic rule that occurs at time t_1, we check a one-second time period $[t_1 - d - 0.5, t_1 - d + 0.5]$ for an occurrence of the same rule, which would constitute a prime. Thus, the model will be able to implicitly estimate the probability of repetition.

Generalized Linear Regression Models (GLMs) can then model the decay by estimating the relationship between d and the probability of rule repetition. The model is designed to predict whether repetition will occur, or, more precisely, whether there is a prime for a given target (priming). Under a no-priming null-hypothesis, we would assume that the priming probability is independent of d. If there is priming, however, increasing d will negatively influence the priming probability (decay). So, we expect a model parameter (DIST) for d that is reliably negative, and lower, if there is more priming.

With this method, we draw multiple samples from the same utterance–for different d, but also for different syntactic rules occurring in those utterances. Because these samples are inter-dependent, we use a grouping variable indicating the source utterance. Because the dataset is sparse with respect to PRIME, balanced sampling is needed to ensure an equal number of data points of priming and non-priming cases (PRIME) is included.

This method has been previously used to confirm priming effects for the general case of syntactic rules by Reitter et al. (2006b). Additionally, the GLM can take into account categorical and continuous covariates that may interact with the priming effect. In the present experiment, we use an interaction term to model the effect of task success.[3] The crucial interaction, in our case, is task success: PATHDEV is the deviation of the paths that the interlocutors drew, normalized to the range [0,1]. The core model is thus PRIME $\sim log(\text{DIST}) * \text{PATHDEV}$.

If IAM is correct, we would expect that the deviation of paths, which indicates negative task success, will negatively correlate with the priming effect.

5.2 Results

Short-term priming reliably correlated (negatively) with the distance, hence we see a decay and priming effect (DIST, $b = -0.151, p < 0.0001$, as shown in previous work).

Notably, path deviation and short-term priming did not correlate. The model showed was no such interaction (DIST:PATHDEV, $p = 0.91$).

We also tested for an interaction with an additional factor indicating whether prime and target were uttered by the same or a different speaker (comprehension-production vs. production-production priming). No such interaction approached reliability (log(DIST):PATHDEV:ROLE, $p = 0.60$).

We also tested whether priming changes over time over the course of each dialogue. It does not. There were no reliable interaction effects of centered prime/target times (log(DIST):log(STARTTIME), $p = 0.75$, log(DIST):PATHDEV:log(STARTTIME), $p = 0.63$). Reducing the model by removing unreliable interactions did not yield any reliable effects.

5.3 Discussion

We have shown that while there is a clear priming effect in the short term, the size of this priming effect does not correlate with task success. There is no reliable interaction with success.

Does this indicate that there is no strong functional component to priming in the dialogue context? There may still be an influence of cognitive load due to speakers working on the task, or an overall disposition for higher priming in task-oriented dialogue: Reitter et al. (2006b) point at stronger priming in such situations. But our results here are difficult to reconcile with the model suggested by Pickering and Garrod (2004), if we take short-term priming as the driving force behind IAM.

Short-term priming decays within a few seconds. Thus, to what extent could syntactic priming help interlocutors align their situation models? In the Map

[3] We use the $A * B$ operator in the model formulas to indicate the inclusion of main effects of the features A and B and their interactions $A : B$.

Task experiments, interlocutors need to refer to landmarks regularly–but not every few seconds. It would be sensible to expect longer-term adaptation (within minutes) to drive dialogue success.

6 Long-term adapation

Long-term adaptation is a form of priming that occurs over minutes and could, therefore, support linguistic and situation model alignment in task-oriented dialogue. IAM and the success of the SVM based method could be based on such an effect instead of short-term priming. Analogous to the the previous experiment, we hypothesize that more adaptation relates to more task success.

6.1 Method

After the initial few seconds, structural repetition shows little decay, but can be demonstrated even minutes or longer after the stimulus. To measure this type of adapation, we need a different strategy to estimate the size of this effect.

While short-term priming can be pin-pointed using the characteristic decay, for long-term priming we need to inspect whole dialogues and construct and contrast dialogues where priming is possible and ones where it is not. Factor SAMEDOC distinguishes the two situations: 1) Priming can happen in contiguous dialogues. We treat the first half of the dialogue as priming period, and the rule instances in the second half as targets. 2) The control case is when priming cannot have taken place, i.e., between unrelated dialogues. Prime period and targets stem from separate randomly sampled dialogue halves that always come from different dialogues.

Thus, our model (PRIME $\sim$ SAMEDOC $*$ PATHDEV) estimates the influence of priming on rule us. From a Bayesian perspective, we would say that the second kind of data (non-priming) allow the model to estimate a prior for rule repetitions. The goal is now to establish a correlation between SAMEDOC and the existence of repetition. If and only if there is long-term adapation would we expect such a correlation.

Analogous to the short-term priming model, we define repetition as the occurrence of a prime within the first document half (PRIME), and sample rule instances from the second document half. To exclude short-term priming effects, we drop a 10-second portion in the middle of the dialogues.

Task success is inverse path deviation PATHDEV as before, which should, under IAM assumptions, interact with the effect estimated for SAMEDOC.

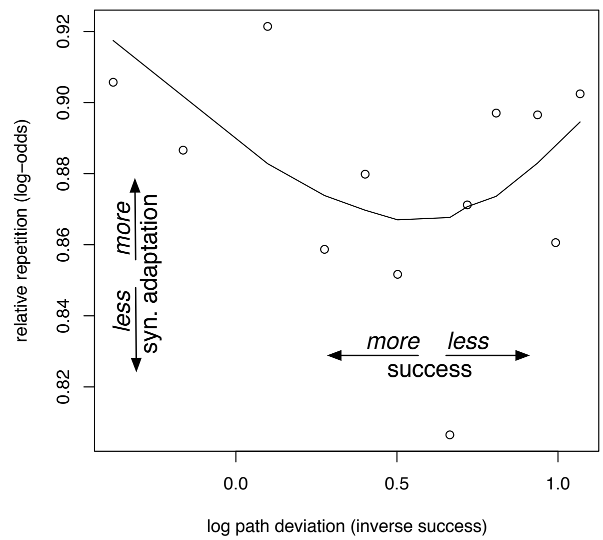

Figure 1: Relative rule repetition probability (chance repetition exluded) over (neg.) task success.

6.2 Results

Long-term repetition showed a positive priming effect (SAMEDOC, $b = 3.303, p < 0.0001$). This generalizes previous experimental priming results in long-term priming.

Long-term-repetition did not interact with (normalized) rule frequency (SAMEDOC:log(RULEFREQ, $b = -0.044, p = 0.35$). The interaction was removed for all other parameters reported.[4]

The effect interacted reliably with the path deviation scores (SAMEDOC:PATHDEV, $b = -0.624, p < 0.05$). We find a reliable correlation of task success and syntactic priming. Stronger path deviations relate to weaker priming.

6.3 Discussion

The more priming we see, the better subjects perform at synchronizing their routes on the maps. This is exactly what one would expect under the assump-

[4]Such an interaction also could not be found in a reduced model with only SAMEDOC and RULEFREQ.

tion of IAM. Also, there is no evidence for stronger long-term adaptation of rare rules, which may point out a qualitative difference to short-term priming.

Of course, this correlation does not necessarily indicate a causal relationship. However, participants in Map Task did not receive an explicit indication about whether they were on the "right track". Mistakes, such as passing a landmark on its East and not on the West side, were made and went unnoticed. Thus, it is not very likely that task success caused alignment to improve at large. We suspect such a possibility, however, for very unsuccessful dialogues. A closer look at the correlation (Figure 1) reveals that while adaptation indeed decreases as task success decreases, adaptation increased again for some of the least successful dialogues. It is possible that here, miscoordination became apparent to the participants, who then tried to switch strategies. Or, simply put: too much alignment (and too little risk-taking) is unhelpful. Further, qualitative, work needs to be done to investigate this hypothesis.

From an applied perspective, the correlation shows that of the repetition effects included in our task-success prediction model, it is long-term syntactic adaptation as opposed to the more automatic short-term priming effect that contributes to prediction accuracy. We take this as an indication to include adaptation rather than just priming in a model of alignment in dialogue.

7 Conclusion

Task success in human-human dialogue is predictable–the more successfully speakers collaborate, the more they show linguistic adaptation. This confirms our initial hypothesis of IAM. In the applied model, knowledge of lexical and syntactic repetition helps to determine task success even after just a few minutes of the conversation.

We suggested two application-oriented tasks (estimating and predicting task success) and an approach to address them. They now provide an opportunity to explore and exploit other linguistic and extra-linguistic parameters.

The second contribution is a closer inspection of structural repetition, which showed that it is long-term adaptation that varies with task success, while short-term priming appears largely autonomous. Long-term adaptation may thus be a strategy that aids dialogue partners in aligning their language and their situation models.

Acknowledgments

The authors would like to thank Frank Keller and the reviewers. The first author is supported by the Edinburgh-Stanford Link.

References

A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC Map Task corpus. *Language and Speech*, 34(4):351–366.

J. Kathryn Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.

J. Kathryn Bock and Zenzi Griffin. 2000. The persistence of structural priming: transient activation or implicit learning? *J of Experimental Psychology: General*, 129:177–192.

H. P. Branigan, M. J. Pickering, and A. A. Cleland. 1999. Syntactic priming in language production: evidence for rapid decay. *Psychonomic Bulletin and Review*, 6(4):635–640.

F. Chang, G. Dell, and K. Bock. 2006. Becoming syntactic. *Psychological Review*, 113(2):234–272.

Kenneth W. Church. 2000. Empirial estimates of adaptation: The chance of two noriegas is closer to $p/2$ than p^2. In *Coling-2000*, Saarbrücken, Germany.

A. Dubey, F. Keller, and P. Sturt. 2005. Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *Proc. HLT/EMNLP-2005*, pp. 827–834. Vancouver, Canada.

Vic Ferreira and Kathryn Bock. 2006. The functions of structural priming. *Language and Cognitive Processes*, 21(7-8).

Stefan Th. Gries. 2005. Syntactic priming: A corpus-based approach. *J of Psycholinguistic Research*, 34(4):365–399.

T. Florian Jäger. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University.

Agnieszka Konopka and J. Kathryn Bock. 2005. Helping syntax out: What do words do? In *Proc. 18th CUNY*. Tucson, AZ.

Martin J. Pickering and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39:633–651.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225.

D. Reitter, J. Hockenmaier, and F. Keller. 2006a. Priming effects in Combinatory Categorial Grammar. In *Proc. EMNLP-2006*, pp.308–316. Sydney, Australia.

D. Reitter, J. D. Moore, and F. Keller. 2006b. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proc. CogSci-2006*, pp. 685–690. Vancouver, Canada.

Benedikt Szmrecsanyi. 2005. Creatures of habit: A corpus-linguistic analysis of persistence in spoken english. *Corpus Linguistics and Linguistic Theory*, 1(1):113–149.

M. Walker, I. Langkilde, J. Wright, A. Gorin, and D. Litman. 2000. Learning to predict problematic situations in a spoken dialogue system: experiments with How may I help you? In *Proc. NAACL-2000*, pp. 210–217. San Francisco, CA.