In Krahmer, E. & Theune, M., Eds., Empiricial Methods in Natural Language Generation, pp. 85–104, 2010, Springer, Berlin/Heidelberg. The original publication is available at springerlink.com

Modelling and Evaluation of Lexical and Syntactic Alignment with a Priming-Based Microplanner

Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp

Sociable Agents Group, CITEC, Bielefeld University
PO-Box 10 01 31, 33501 Bielefeld, Germany
{hbuschme, kbergman, skopp}@TechFak.Uni-Bielefeld.DE

Abstract. Alignment of interlocutors is a well known psycholinguistic phenomenon of great relevance for dialogue systems in general and natural language generation in particular. In this chapter, we present the alignment-capable microplanner SPUD *prime*. Using a priming-based model of interactive alignment, it is flexible enough to model the alignment behaviour of human speakers to a high degree. We demonstrate that SPUD *prime* can account for lexical as well as syntactic alignment and present an evaluation on corpora of task-oriented dialogue that were collected in two experiments designed to investigate the alignment behaviour of humans in a controlled fashion. This will allow for further investigation of which parameters are important to model alignment and how the human–computer interaction changes when the computer aligns to its users.

Keywords: interactive alignment model, lexical and syntactic alignment, adaptation, microplanning

1 Introduction

A well known phenomenon in dialogue situations is *alignment* of the interlocutors. An illustrative example is given by Levelt and Kelter [17], who telephoned shops and either asked the question "What time does your shop close?" or the question *"At* what time does your shop close?". The answers were likely to mirror the form of the question. When asked "At what ...?", answers tended to begin with the preposition 'at' (e.g., "At five o'clock."). Conversely, when asked "What ...?", answers tended to begin without the preposition (e.g., "Five o'clock."). Similar alignment phenomena can be observed in many aspects of speech production *inter alia* in syntactic and lexical choice.

Pickering and Garrod [19] present the *interactive alignment model* bringing together all alignment phenomena of speech processing in dialogue. According to this model, human language comprehension and production are greatly facilitated by alignment of the interlocutors during conversation. The process of alignment is explained through mutual priming of the interlocutors' linguistic representations. Thus, it is automatic, efficient, and non-conscious. A stronger claim of the authors is that alignment — in combination with routines and a dialogue lexicon — is a prerequisite for fluent speech production in humans.

Alignment effects also occur in human–computer interaction. Brennan [7] and Branigan et al. [6] present evidence that syntactic structures and lexical items used by a computer are subsequently adopted by users. For this reason, alignment is an important concept for natural language human–computer interaction in general, and for dialogue systems with natural language generation in particular. Integrating ideas from the interactive alignment model into the microplanning component of natural language generation systems should be beneficial for several reasons. First, microplanning may become more efficient since the subsets of rules or lexical items in the dialogue lexicon that have been used before can be preferentially searched. Second, due to *self*-alignment, the output of the system can become more consistent and thus easier to understand for the user. Finally, mutual alignment of user and dialogue system might make the conversation itself more natural and, presumably, cognitively more lightweight for the user.

In this chapter we present a computational model for parts of the interactive alignment model that are particularly important in the context of natural language generation. We describe how this model has been incorporated into the existing SPUD *lite* system [23, 22] to yield the *alignment-capable* microplanner SPUD *prime*. In Sect. 2 we describe previous approaches to integrate alignment into natural language generation. In Sects. 3 and 4, we present our priming-based model of alignment and its implementation in SPUD *prime*. In Sect. 5, we demonstrate that SPUD *prime* works as specified and describe and discuss the results of an empirical evaluation study on two corpora of task-oriented dialogue. In Sect. 6 we discuss our work and in Sect. 7 we conclude and describe possible future directions.

2 Related Work

Computational modelling is an important methodology for evaluating and testing psycholinguistic theories. Thus, it is certainly not a new idea to implement the interactive alignment model computationally. Indeed, a call for "explicit computational models" is made as early as in the open peer commentary on Pickering and Garrod's paper [19].

Brockmann et al. [9] and Isard et al. [13] present a 'massive over-generation' approach to modelling alignment and individuality in natural language generation. Their system generates a huge number of alternative sentences — up to 3000 — and evaluates each of these sentences with a trigram model consisting of two parts: a default language model computed from a large corpus and a cache model which is calculated from the user's last utterance. The default language model is linearly interpolated with the cache model, whose influence on the resulting combined language model is determined by a weighting factor $\lambda \in [0, 1]$ that controls the amount of alignment the system exhibits.

Purver et al. [20] take a more formal approach. They use an implementation of the Dynamic Syntax formalism, which uses the same representations and mechanisms for parsing as well as for generation of natural language, and extend it with a model of context. In their model, context consists of two distinct representations: a record of the semantic trees generated and parsed so far and a record of the transformation actions used for the construction of these semantic trees. *Re-use* of semantic trees and actions is used to model many dialogue phenomena in Dynamic Syntax and can also explain alignment. Thus, the authors declare alignment to be a corollary of context re-use. In particular, re-use of actions is assumed to have a considerable influence on alignment in natural language generation. Instead of looking through the complete lexicon each time a lexical item is chosen, this kind of lexical search is only necessary if no action — which constructed the same meaning in the given context before — exists in the record. If such an action exists, it can simply be re-used, which obviously leads to alignment.

A completely different approach to alignment in natural language generation is presented by de Jong et al. [15], whose goal is to make a virtual guide more believable by aligning to the user's level of politeness and formality. In order to achieve this, the virtual guide analyses several features of the user's utterance and generates a reply with the same level of politeness and formality. According to the authors, lexical and syntactic alignment occur automatically because the lexical items and syntactic constructions to choose from are constrained by the linguistic style adopted.

Finally, Bateman [1] advocates another proposal according to which alignment in dialogue is predictable because communication is an inherently social activity. Following the social-semiotic view of language, Bateman suggests to model alignment as arising from register and micro-register. More specifically, in his opinion priming of a linguistic representation is comparable with preselecting a micro-register that must be considered when generating an utterance in a particular social context.

The approaches presented above primarily focus on the linguistic and social aspects of alignment in natural language generation. The work of Brockmann et al. [9] and Isard et al. [13] concentrates on the surface form of language, Bateman [1] sees alignment arising from social-semiotic aspects, and Purver et al. [20] are primarily interested in fitting alignment into a formal linguistic framework. In this paper we adopt a more psycholinguistic and cognitive stance on alignment. Pickering and Garrod [19] suggest that low-level priming is the basic mechanism underlying interactive alignment. Here, we propose that computational modelling of these priming mechanisms also opens up an interesting and new perspective for alignment in natural language generation.

3 A Priming-based Model of Alignment

We are interested here in those parts of the interactive alignment model that are most relevant for microplanning in natural language generation and it is out

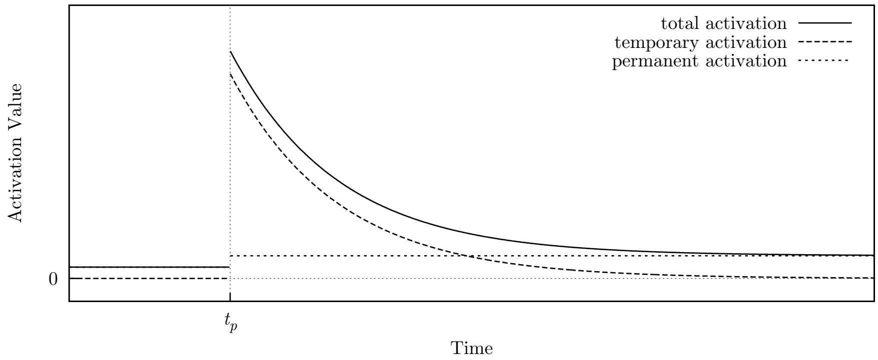

Fig. 1. Change of activation values of a linguistic structure primed at the point of time t_p. In this example, the total activation value is simply the sum of the temporary and the permanent activation values.

of our scope to model all the facets and details of direct/repetition priming in the alignment of linguistic representations. Exact timing effects, for instance, are likely to be not very relevant as, in an actual system, it does not matter how many milliseconds faster the retrieval of a primed lexical item is in contrast to the retrieval of an item that is not primed. For this reason we adopt an idealised view, in which priming of linguistic structures results from two basic activation mechanisms:

Temporary activation This kind of activation should increase abruptly and then decrease slowly over time until it reaches zero again.

Permanent activation This kind of activation should increase by a certain quantity and then maintain the new level.

Figure 1 shows how the different activation values should change over time when primed at the point of time t_p.

The two mechanisms of priming are in accordance with empirical findings. Branigan et al. [5] present evidence for rapid decay of activation of primed syntactic structures, whereas Bock and Griffin [4] report evidence for their long(er) term activation. In any case, Reitter [21] found both types of priming in his analysis of several corpora, with temporary activation being the more important one. The assumption that both mechanisms play a role in dialogue is also supported by Brennan and Clark [8] whose terminology will be followed in this paper: temporary priming will be called 'recency of use effects' and permanent priming will be called 'frequency of use effects'.

Reitter [21] assumes the repetition probability of primed syntactic structures to depend logarithmically on the distance between priming and usage. Here, we model recency of use effects by a more general *exponential decay* function,

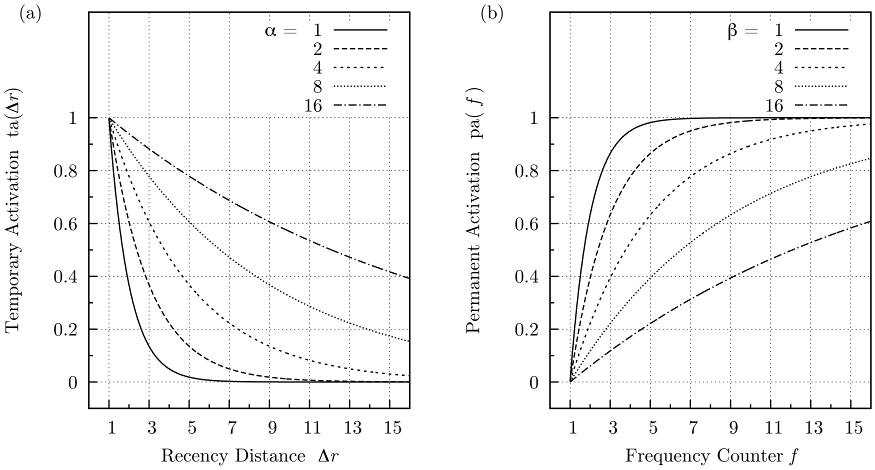

Fig. 2. Plots of the mathematical functions that model recency and frequency effects. Plot (a) displays temporary activation depending on the recency of priming. Plot (b) shows permanent activation depending on the frequency count. Both are shown for different values of the slope parameter α respectively β.

modified to meet the needs for modelling activation decay of primed structures:

$$ta(\Delta r) = \exp\left(-\frac{\Delta r - 1}{\alpha}\right), \tag{1}$$

$$\Delta r \in \mathbb{N}^+;\ \alpha > 0; \quad ta \in [0, 1]$$

$ta(\Delta r)$ is the temporary activation value of a linguistic structure depending on the distance Δr between the current time T and the time r at which the structure was primed. The slope of the function is determined by the parameter α. Additionally, the function is shifted right in order to yield an activation value of 1 for $\Delta r = 1$. This shift is due to the assumption of discrete time steps with a minimal distance of 1. A plot of $ta(\Delta r)$ with different values for α is given in Fig. 2a.

Using exponential decay to model temporary activation appears to be a sensible choice that is often used to model natural processes. The advantage of this model of temporary activation lies in its flexibility. By changing the slope parameter α, different empirical findings as well as variation among humans can be modelled easily.

Next, a mathematical model for frequency of use effects is needed. To prevent that frequency effects lead to an ever increasing activation value, a maximum activation level exists. This is also found in Reitter's corpus studies [21], which indicate that the frequency effect is inversely connected to the recency effect. Here, we model frequency effects with a general *exponential saturation* function, modified to meet the requirements for modelling permanent activation of linguistic structures:

$$pa(f) = 1 - \exp\left(-\frac{f-1}{\beta}\right), \tag{2}$$

$$f \in \mathbb{N}^+;\ \beta > 0; \quad pa \in [0,1]$$

The most important point to note here is that the permanent activation value $pa(f)$ is not a function of time but a function of the frequency-counter f attached to each linguistic structure. Whenever a structure is primed, its counter is increased by the value of 1. Again, the slope of the function is determined by the parameter β and the function is shifted right in order to get an activation value of 0 for $f = 1$. A plot of $pa(f)$ with different slope parameters is given in Fig. 2b. Similar to the advantages of the model of temporary activation, this model for frequency effects is very flexible so that different empirical findings and human individuality can be expressed easily.

Now, both priming models need to be combined for a model of alignment. We opted for a weighted linear combination of temporary and permanent activation:

$$ca(\Delta r, f) = \nu \cdot ta(\Delta r) + (1-\nu) \cdot pa(f), \tag{3}$$

$$0 \le \nu \le 1; \quad ca \in [0,1]$$

Different values of ν allow different forms of alignment. With a value of $\nu = 0.5$ recency and frequency effects are equally important, with a value of $\nu = 1$ alignment depends on recency only, and with a value of $\nu = 0$ alignment is governed solely by frequency. Being able to adjust the influence of the different sorts of priming on alignment is crucial as it has not yet been empirically determined to what extent recency and frequency of use affect alignment (in Sects. 5.3 and 5.4 we will exploit this flexibility for matching empirical data).

In contrast to the models of alignment presented in Sect. 2, the computational alignment model presented here will not only consider alignment between the interlocutors (interpersonal- or *other-alignment),* but also alignment to oneself (intrapersonal- or *self-alignment*). Pickering et al. [18] present results from three experiments which suggest self-alignment to be even more important than other-alignment. In our model, self-alignment is accounted for with the same priming-based mechanisms. To this end, four counters are attached to each linguistic structure:

- Δr_s: recency of use by the system itself
- Δr_o: recency of use by the interlocutor
- f_s: frequency of use by the system itself
- f_o: frequency of use by the interlocutor

The overall activation value of the structure is a linear combination of the combined activation value $ca(\Delta r_s, f_s)$ and the combined activation value $ca(\Delta r_o, f_o)$ from equation (3):

$$act(\Delta r_s, f_s, \Delta r_o, f_o) = \lambda \cdot \big(\mu \cdot ca(\Delta r_s, f_s) + (1-\mu) \cdot ca(\Delta r_o, f_o)\big), \tag{4}$$

$$0 \leq \lambda, \mu \leq 1; \quad act \in [0, 1]$$

Again, by changing the factor μ, smooth interpolation between pure self-alignment ($\mu = 1$) and pure other-alignment ($\mu = 0$) is possible, which can account for different empirical findings or human individual differences. Furthermore, the strength of alignment is modelled with a scaling factor λ, which determines whether alignment is considered during generation ($\lambda > 0$) or not ($\lambda = 0$).

4 The Alignment-Capable Microplanner SPUD *prime*

The previously described priming-based model of alignment has been implemented by extending the integrated microplanning system SPUD *lite* [22]. SPUD *lite* is a lightweight Prolog re-implementation of the SPUD microplanning system [23] based on the context-free tree rewriting grammar formalism TAGLET. Not only the microplanner itself, but also the linguistic structures (the initial TAGLET trees) are represented as Prolog clauses.

SPUD *lite* carries out the different microplanning tasks (lexical choice, syntactic choice, referring expression generation and aggregation) at once by treating microplanning as a search problem. During generation it tries to find an utterance that is in accordance with the constraints set by its input (a grammar, a knowledge base and a query). This is done by searching the search space spanned by the linguistic grammar rules and the knowledge base until a goal state is found. Non-goal search states are preliminary utterances that are extended by one linguistic structure in each step until a syntactically complete utterance is found which conveys all the specified communicative goals. Since this search space is large even for relatively small grammars, a heuristic greedy search strategy is utilised.

Our alignment-capable microplanner SPUD *prime* extends SPUD *lite* in several ways. First, we altered the predicate for the initial TAGLET trees by adding a unique identifier ID as well as counters for self/other-recency/frequency values (r_s, f_s, r_o and f_o; see Sect. 3). The activation value of an initial tree is then calculated with equation (4).

Furthermore, we have created a mechanism that enables SPUD *lite* to change the recency and frequency information attached to the initial trees on-line during generation. This is done in three steps with the help of Prolog's meta-programming capabilities: First, the clause of a tree is retrieved from the knowledge base. Second, it is retracted from the knowledge base. Finally, the clause is (re-)asserted in the knowledge base with updated recency and frequency information. As a welcome side effect of this procedure, primed initial trees are moved to the top of the knowledge base and — since Prolog evaluates clauses and facts in the order of their appearance in the knowledge base — they can be accessed earlier than unprimed initial trees or initial trees that were primed longer ago. Thus, in SPUD *prime* recency of priming directly influences the access of linguistic structures.

Most important, the activation values of the initial trees are considered during generation. Thus, in addition to the evaluation measures used by SPUD *lite*'s heuristic state evaluation function, the mean activation value

$$\overline{act}(S) = \frac{\sum_{i=1}^{N} act_{t_i}(\Delta r_{s_{t_i}}, f_{s_{t_i}}, \Delta r_{o_{t_i}}, f_{o_{t_i}})}{N}$$

of the N initial trees $\{t_1, \ldots, t_N\}$ of a given search state S is taken into account as a further evaluation measure. Hence, when SPUD *prime* evaluates (otherwise equal) successor search states, the one with the highest mean activation value is chosen as the next current state.

5 Evaluation

In order to show that our priming-based alignment model and its implementation work as intended, we first demonstrate that SPUD *prime* is in principle capable of lexical and syntactic alignment as well as that it can display recency and frequency of use effects by simulating some — admittedly rather artificial — interactions (Sect. 5.1). Having established these abilities, we then evaluate SPUD *prime* empirically on two corpora collected in two psycholinguistic experiments designed to investigate the alignment behaviour of humans in a controlled fashion (Sects. 5.2–5.4).

5.1 Demonstrating Lexical and Syntactic Alignment

A simple demonstration that SPUD *prime* displays alignment phenomena is to do tests that resemble the course of psychological experiments (e.g., Bock [3]) where subjects are primed and the influence of the prime is observed in their verbal behaviour. This can be done in four steps for, e.g., lexical alignment:

1. Querying SPUD *prime* to generate an utterance u_1 from the communicative goals CG: Utterance u_1 uses lexical item l_1.
2. Priming a lexical item l_2 that is synonymous to lexical item l_1.
3. Querying SPUD *prime* to generate an utterance u_2 from the same communicative goals CG.
4. Analysing utterance u_2: if it uses the primed lexical item s_2, then SPUD *prime* displays lexical alignment, otherwise it does not.

In the following we use this and similar tests in order to demonstrate that SPUD *prime* displays lexical alignment, syntactic alignment as well as recency and frequency of use effects. The steps above are translated into commands for SPUD *prime* (used together with a small German TAGLET grammar in a landmark description domain). The model is set up with the parameters $\alpha = 2$, $\beta = 16$, $\lambda = 1$, $\mu = 0.6$ and $\nu = 0.8$ — weighting self-alignment stronger than other-alignment and recency effects stronger than frequency effects. Anyway, the parameter setting is not too important in this demonstration as we just want to show that SPUD *prime* is in principle capable of displaying alignment phenomena.

Lexical Alignment. SPUD *prime*'s ability to display lexical alignment can be demonstrated by following steps 1–4 directly.

(1a) The following knowledge base is loaded:

```
shared(entity(church-1, single)).
shared(instance_of(church-1, church)).
private(entity(window-7, single)).
private(instance_of(window-7, window)).
private(property(window-7, round)).
private(part_of(church-1, window-7)).
```

(it states that there exists a church `church-1` that has a round window `window-7`).

(1b) SPUD *prime* is requested to generate an utterance that communicates the structure of `church-1`:

```
spudprime(initial_state(structure(church-1), _,
    [part_of(church-1, window-7),
     entity(window-7, single),
     instance_of(window-7, window),
     property(window-7, round)]), W).
```

For this request, SPUD *prime* generates the output *'Die Kirche hat ein rundes Fenster.'* ('The church has got a round window.').

(1c) Now it is pretended that an interlocutor uses the (more or less) synonymous lexical item *'kreisförmig'* ('circular') instead of *'rund'* ('round'). The initial tree of the lemma *'kreisförmig'* (`dynlex-502`) is therefore primed with the following SPUD *prime* query:

```
sp_fake_interlocutor_rule_usage(dynlex-502).
```

(1d) SPUD *prime* is requested to regenerate the utterance with the same query used in (1b). This time the output *'Die Kirche hat ein kreisförmiges Fenster.'* ('The church has got a circular window.') is generated.

To conclude, after priming rule `dynlex-502` the corresponding lexical item in the utterance changes as the model predicts. In (1b) SPUD *prime* uses the word *'rund'* (because it happens to be easier to access in the knowledge base), in (1d) it uses the synonymous word *'kreisförmig'* (because it has a higher activation). Hence, SPUD *prime* displays lexical alignment.

Syntactic Alignment. Since lexicon and syntax are represented uniformly in TAGLET, the steps to demonstrate syntactic alignment are the same as for lexical alignment.

(2a) The following knowledge base is loaded:

```
shared(entity(church-1, single)).
shared(instance_of(church-1, church)).
private(relpos(church-1, left)).
```

(it states that there exists a church `church-1` that is on the left side).

(2b) SPUD *prime* is requested to generate an utterance that communicates the position of `kirche-1`:

```
spudprime(initial_state(position([church-1]), _,
    [relpos(church-1, left)]), W).
```

For this request, SPUD *prime* generates the output *'Die Kirche ist auf der linken Seite.'* ('The church is on the left side.')

(2c) Now it is pretended that an interlocutor uses a different syntactic construction. The initial tree of that construction (`rule-522`) is therefore primed with the following SPUD *prime* query:

```
sp_fake_interlocutor_rule_usage(rule-522).
```

(2d) Lastly, SPUD *prime* is requested to regenerate the utterance with the same query used in (2b). This time the output *'Auf der linken Seite ist die Kirche.'* ('On the left side is the church.') is generated.

To conclude, after priming `rule-522`, the syntactic structure of the utterance changes as the model predicts. In (2b) SPUD *prime* uses initial tree `rule-526` to generate the utterance (again, because it happens to be easier to access in the knowledge base), in (2d) it uses the primed initial tree `rule-522` (again, because it has a higher activation). Hence, SPUD *prime* displays syntactic alignment.

Recency and Frequency Effects. In the two previous tests, priming with `sp_fake_interlocutor_rule_usage/1` changes both the recency and the frequency information that is attached to initial trees (cf. Sects. 3 and 4). However, the interesting aspect of recency and frequency of use is its behaviour over time, which we test here. To simplify matters, this test is based on the first one.

(3a) See (1a).
(3b) See (1b). Additionally, it is assumed that the self-frequency f_s of `dynlex-9` (*'rund'*) has the value 10 instead of 1. This can be set with the SPUD *prime* query

```
sp_set_frecency(dynlex-9, 1, 10, 1, 1).
```

which sets a given rule's four counters to new values (*frecency* is short for *'fre*quency and re*cency'*). In this case Δr_s is set to 1, f_s to 10 and Δr_o and f_o to 1. As all counters defaulted to 1, only f_s changed.

(3c) See (1c).
(3d) See (1d).

(3e) Now it is pretended that some time goes by. This can be set with the query

```
sp_increase_recency_counter(10).
```

which increases the current point of time T by a value of 10.

(3f) Finally, SPUD *prime* is requested to regenerate the utterance with the same query used in (3d). This time the output *'Die Kirche hat ein rundes Fenster.'* ('The church has got a round window.') is generated again.

To conclude, similar to test 1, the lexical item *'kreisförmig'* primed in (3c) is used in the utterance generated in (3d) — although the lexical item *'rund'* has a frequency value five times as high. This demonstrates that recency is more important than frequency of use (given the chosen parameters we use here). After a short period of time (3e) SPUD *prime* uses the word *'rund'* again in (3f): the temporary activation based on the recency value has decayed and permanent activation based on the frequency value is higher again. Hence, SPUD *prime* displays both, recency and frequency of use effects.

The three tests show that SPUD *prime* displays the alignment phenomena predicted by our priming based model of alignment: syntactic and lexical alignment as well as recency and frequency of use effects.

5.2 Empirical Evaluation Method

For the empirical evaluation of our priming-based model of alignment and its implementation in SPUD *prime* we use two small corpora of recorded and transcribed spoken dialogues between human interlocutors. These were collected in two psycholinguistic experiments designed to investigate the alignment behaviour of humans in a controlled fashion. The participants' task was to play the 'Jigsaw Map Game', in which different objects have to be placed correctly on a table. Each participant has a unique set of cards and a box of objects and they take turns in giving each other instructions of how to place the next object in relation to the objects that are already on the table (cf. Weiß et al. [24, 25]).

In our evaluation, we concentrate on the generation of the object names (i.e., nouns), by simulating their usage in the dialogues. In each simulation run, SPUD *prime* plays the role of one of the two speakers interacting with a simulated interlocutor who behaves exactly as in the real experiments. With this test setup we examined, first, how well SPUD *prime* can model the alignment behaviour of a real speaker in a real dialogue context and, second, whether our model is flexible enough to consistently emulate different speakers with different alignment behaviour.

In order to find the best model, i.e., the best point $(\alpha, \beta, \mu, \nu)$ in parameter space, for each speaker, we simulated all tests with all parameter combinations and counted the number of mismatches between our model's choice and the real speaker's choice. To make this exhaustive search possible, we limit the set of values for the parameters α and β to $\{1, 2, 4, 6, 8, 10, 14, 18, 24, 30\}$ and the set of values for the parameters μ and ν to $\{0, 0.1, 0.2, \ldots, 1\}$, resulting in a total of $11^2 \times 10^2 = 12100$ points in parameter space. Since we want to investigate alignment, λ is constantly set to 1.

In the next section (5.3) we describe the evaluation on the first corpus and give an example of how it is done. Thereafter we describe the evaluation on the second corpus (Sect. 5.4).

5.3 Corpus 1: Learning of Referring Nouns

The first corpus that we used consists of eight recorded and transcribed dialogues between pairs of two interlocutors — named (A) and (B) — that play the 'Jigsaw Map Game'. Each speaker learned[1] a set of object names before playing the game, such that both use the same names for all but three objects[2]. Due to this precondition, both speakers use the same lexical referring expressions for most objects and the speaker's lexical alignment behaviour for the differently named objects can be observed easily. The experiment is described in further detail in Weiß et al. [24].

To illustrate our evaluation method, we first present and discuss the simulation of one particular dialogue (number 7) from the corpus from the perspective of participant (A). Both interlocutors learned the object names *'Raute'* ('rhombus'), *'Ring'* ('ring'), *'Schraube'* ('bolt') and *'Würfel'* ('dice'), additionally participant (A) learned *'Spielfigur'* ('token'), *'Ball'* ('sphere') and *'Block'* ('cuboid') and participant (B) learned *'Männchen'* ('token'), *'Kugel'* ('sphere') and *'Klotz'* ('cuboid'). In our simulation, we focus on the use of the differently learned names (the *targets*) and not on the other names (the *non-targets*). Table 1 shows the sequence of target nouns as they occurred in one of the real dialogues (non-targets omitted).

For each point in parameter space $(\alpha, \beta, \mu, \nu)$ the dialogue is simulated in the following way:

- When participant (A) referred to a *target* object in the dialogue, SPUD *prime* is queried to generate a noun for the target object and the corresponding rule(s) are primed automatically. Then it is recorded whether the noun actually generated is the noun used in the actual dialogue (match) or not (mismatch).
- When participant (A) used a *non-target* object name in the dialogue, self-priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated (i.e., the recency and frequency counters are increased).

[1] The participants had to learn the object names in the following way: First, a list of the objects and their names was presented to them. Second, after reading the task instructions, the same list was shown to them again. Finally, they had to demonstrate that they memorised the names by naming the objects twice, in a written 'test', and verbally to the experimenter.

[2] Note, however, that the participants were not explicitly instructed to use the learned object names during the experiment.

Table 1. Sequence of referring target nouns used by participants (A) and (B) in our example dialogue 7.

	B:	der Klotz		B:	der Ball			der Klotz
1	A:	die Spielfigur	11	A:	der Ball	18	A:	das Männchen
2		der Klotz	12		der Ball	19		der Klotz
	B:	das Männchen		B:	die Kugel		B:	das Männchen
		der Klotz			das Männchen	20	A:	der Ball
3	A:	die Spielfigur	13	A:	der Ball	21	A:	das Männchen
	B:	das Männchen		B:	die Kugel		B:	der Ball
4	A:	das Männchen	14	A:	der Klotz			das Männchen
5		das Männchen	**15**	**A:**	**die Kugel**	**22**	**A:**	**die Kugel**
6		das Männchen	16		der Klotz	23	A:	der Ball
7		das Männchen		B:	der Klotz		B:	der Klotz
8		das Männchen			die Kugel	24	A:	der Ball
	B:	das Männchen			der Klotz		B:	der Klotz
9	A:	das Männchen	17	A:	der Klotz	25	A:	der Klotz
10		der Ball		B:	das Männchen			———

Table 2. Number of points in parameter space p leading to m mismatches for participant (A) in dialogue 7.

No. of Mismatches (m)	0	1	2	3	4	5	6	7	8	9	10	...
Points in par. space (p)	0	0	4	833	3777	2248	3204	1105	478	148	294	0

– When participant (B) used an object name (target or non-target), priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated.

The evaluation measure for a specific point in parameter space is the number of mismatches it produces when simulating a dialogue. Thus the point (or rather points) in parameter space that produce the least number of mismatches are the ones that best model the particular speaker under consideration. For participant (A) of our example dialogue the distribution of points in parameter space p producing m mismatches is shown in Table 2. Four points in parameter space produce only two mismatches (in phrase 15 and 22; cf. Table 1) and thus our priming-based alignment model can account for 92% of the target nouns produced by speaker (A). However, it must be noted that these two mismatches occur at points in the dialogue where the alignment behaviour of (A) is not straightforward. At target noun 15, both interlocutors have already used the name *'Ball'* and then both switch to *'Kugel'*. The mismatch at target 22 is a special case: (A) used *'Kugel'* and immediately corrected himself to *'Ball'*, the name he learned prior to the experiment. In this case it seems as if (A) suddenly remembers the learning phase before and after the task instructions.

We simulated the noun production for each of the interlocutors from the first corpus. One dialogue has been excluded from the data analysis as the dialogue partners used nouns that none of them had learned in the priming phase. For each of the remaining 14 interlocutors we varied the parameters α, β, μ and ν as described above to identify those point(s) in parameter space that result in the least number of mismatches.

Each interlocutor produced between 18 and 32 target nouns (N = 14, M = 23.1, SD = 3.9). Our simulation runs contain between 0 and 19 mismatches overall (N = 169400, M = 6.4, SD = 3.4). The minimal number of mismatches for each speaker simulation ranges between 0 and 6 (N = 14, M = 2.3, SD = 1.7). That is, our model can simulate a mean of 89.8% of all target nouns (N = 14, Min = 66.7%, Max = 100.0%, SD = 8.2%), which is an improvement of 24.6% on the baseline condition (alignment switched off), where 65.3% of the target nouns are generated correctly (N = 14, Min = 36.0%, Max = 100.0%, SD = 7.1%). As already illustrated in the example simulation, mismatches typically occur at points in the dialogue where the alignment behaviour of the human interlocutor is not straightforward.

As displayed in Table 3 the parameter assignments resulting in least mismatches differ considerably from speaker to speaker. However, there are some remarkable trends to be observed in the data. As concerns the parameter μ, which determines the combination of self- and other-alignment, the majority of values are in the upper range of the interval [0,1]. For 8 of 14 speakers the mean is above 0.7 with relatively low standard deviations. Only for one speaker (P13) the mean μ is below 0.3. The overall mean value of μ is 0.666 (N = 14, SD = 0.206). Thus, the parameter values indicate a considerable tendency toward self-alignment in contrast to other-alignment.

For the parameter ν, which interpolates between recency and frequency effects of priming, the results are less revealing. For two speaker simulations (P13 and P48) the mean ν is 0.166 or lower, for another four speaker simulations the mean ν is above 0.7. That is, our model produces good matching behaviour in adopting different alignment strategies, depending either primarily on frequency or recency, respectively. All other simulations, however, are characterised by a mean ν in the medium range along with a relatively high standard deviation. The mean ν of all speakers is 0.560 (N = 14, SD = 0.274).

One shortcoming of the first experiment and corpus is that participants explicitly learned the object names prior to the game, which is a clear difference from the alignment effects (acquisition of object names by lexical priming) that occur during the game itself. The sudden remembrance of the learning phase (this object is called *'Ball'*) mentioned in the example above might be one consequence of this. Furthermore, it could not be controlled how eager the participants were in learning the names, so it is not clear how the rules in SPUD *prime*'s knowledge base should be initialised. Our — somewhat arbitrary — decision was to prime the rules (i.e., increase their recency and frequency counters prior to the simulation) for the learned object names three times.

5.4 Corpus 2: Implicit Acquisition of Referring Nouns

In order to overcome the shortcomings just mentioned, a second and slightly different study was conducted. The corpus collected in this experiment consists of

Table 3. Mean parameter values for those simulation runs that result in a minimal number of mismatches for each speaker of the first corpus (T = number of targets, m = least number of mismatches, % = percentage of targets that could be simulated, # p = number of points in parameter space that lead to m mismatches).

					α		β		μ		ν	
	T	**m**	**%**	**# p**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
VP13	25	2	92.0	4	3.00	1.16	19.50	9.15	0.300	0.000	0.100	0.000
VP14	19	1	94.7	72	5.53	1.52	14.32	9.61	0.819	0.040	0.901	0.108
VP17	25	1	96.0	200	1.66	0.82	12.94	9.53	0.353	0.169	0.955	0.069
VP18	22	3	86.4	2445	15.37	8.76	10.98	9.76	0.597	0.211	0.706	0.236
VP19	22	0	100.0	4321	11.81	9.49	11.01	8.93	0.824	0.148	0.388	0.291
VP20	18	2	88.9	8	1.00	0.00	15.75	9.29	0.738	0.052	0.388	0.146
VP23	18	6	66.7	987	6.85	6.68	12.08	9.35	0.331	0.374	0.400	0.330
VP24	29	3	89.7	256	12.95	9.70	13.63	8.94	0.538	0.201	0.468	0.298
VP39	32	5	84.4	1	1.00	0.00	2.00	0.00	0.900	0.000	0.800	0.000
VP40	26	0	100.0	3504	12.08	9.33	10.30	8.75	0.843	0.147	0.343	0.282
VP41	21	2	90.5	609	11.37	8.48	15.34	8.92	0.770	0.106	0.655	0.213
VP42	22	3	86.4	30	6.00	1.49	17.53	9.02	0.783	0.059	0.760	0.122
VP47	20	2	90.0	326	13.75	7.79	13.53	9.51	0.772	0.095	0.816	0.166
VP48	24	2	91.7	2478	12.87	9.55	10.74	8.54	0.764	0.175	0.166	0.148
M	23.1	2.3	89.8	1089	8.23	5.75	12.83	9.18	0.666	0.137	0.560	0.185
SD	4.1	1.7	8.2	1468	5.20	4.01	4.12	0.37	0.206	0.097	0.274	0.099

12 interactions, each divided into two parts [25]. As before, participants played the 'Jigsaw Map Game', but this time they did not learn the object names explicitly. In the first part of each interaction (the *priming phase),* a naïve participant (A) played the game with a confederate (C) that was instructed to use specific object names so that (A) could acquire them implicitly through lexical priming. In the second part of each interaction (the *usage phase),* participant (A) then played the game with a second naïve participant (B).

For the evaluation on this corpus, we, again, simulated the dialogues for each point in parameter space $(\alpha, \beta, \mu, \nu)$, first simulating the priming phase followed by a simulation of the usage phase. In the priming phase the dialogue was simulated in the following way:

- When the confederate (C) used an object name (target or non-target) in the dialogue, priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated (i.e., the recency and frequency counters are increased).
- When participant (A) used an object name (target or non-target), self-priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated.

Grounded in the knowledge base from the priming phase, the usage phase dialogue between participants (A) and (B) was simulated — from the perspective of (A) only — in the same way described in Sect. 5.3.

Table 4. Mean parameter values for those simulation runs that result in a minimal number of mismatches for each speaker (A) in the usage phase of the second corpus (T = number of targets, m = least number of mismatches, % = percentage of targets that could be simulated, # p = number of points in parameter space that lead to m mismatches).

					α		β		μ		ν	
	T	**m**	**%**	**# p**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
V1	16	3	81.3	1697	18.06	8.11	13.75	9.80	0.221	0.163	0.676	0.252
V4	9	3	66.7	477	13.20	9.69	12.58	8.38	0.096	0.101	0.100	0.093
V6	13	1	92.3	2967	9.66	7.70	14.10	9.51	0.498	0.179	0.772	0.193
V7	16	7	56.3	6221	12.54	9.35	10.62	8.80	0.678	0.271	0.335	0.282
V8	16	3	81.3	417	18.06	6.88	14.25	9.87	0.114	0.102	0.851	0.120
V19	11	1	90.9	119	5.33	1.67	11.77	9.51	0.761	0.063	0.707	0.136
V33	13	2	84.6	151	24.04	5.58	14.72	9.09	0.313	0.110	0.637	0.195
V34	14	3	78.6	39	24.41	4.50	16.41	0.06	0.621	0.047	0.933	0.081
V35	17	1	94.1	1582	14.38	7.96	12.91	9.51	0.286	0.234	0.663	0.206
V36	11	0	100.0	928	7.63	3.24	11.90	9.40	0.319	0.235	0.619	0.184
V37	12	3	75.0	2774	12.98	8.28	13.69	9.57	0.276	0.215	0.781	0.200
V38	22	4	81.8	2118	7.12	7.32	11.97	9.28	0.193	0.158	0.547	0.211
M	14.2	2.6	81.9	1624	13.95	6.69	13.22	9.40	0.365	0.156	0.635	0.179
SD	3.5	1.8	12.2	1777	6.24	2.46	1.58	0.47	0.222	0.073	0.227	0.061

In the usage phase, participants (A) produced between 9 and 22 target nouns ($N = 12$, $M = 14.2$, SD = 3.5). Our simulation runs contain between 0 and 15 mismatches overall ($N = 145200$, $M = 5.9$, SD = 3.5). The minimal number of mismatches for each speaker simulation ranges between 0 and 7 ($N = 12$, $M = 2.6$, SD = 1.8). That is, our model can simulate a mean of 81.9% of all target nouns ($N = 12$, Min = 56.3%, Max = 100.0%, SD = 12.2%), which is an improvement of 17.3% on the baseline condition (alignment switched off), where 64.3% of the target nouns are generated correctly ($N = 12$, Min = 22.2%, Max = 92.3%, SD = 21.9%).

As in the first corpus evaluation, the parameter assignments resulting in least mismatches differ considerably from speaker to speaker (Table 4). Furthermore, comparing the results of this evaluation to the results of the first (Table 3) reveals further similarities. There is no significant difference in the mean least number of mismatches and the mean coverage between the two evaluation studies (2.3 in the first versus 2.6 in the second and 89.8% versus 81.9%). There is also no significant difference in the number of points in parameter space that lead to the least number of mismatches (1089 versus 1624) in the values β (means are 12.83 versus 13.22) and ν (means are 0.560 versus 0.635).

One remarkable difference, however, can be observed between the two μ values which control the relation of self- and other-alignment. Their means for the simulation runs with least number of mismatches for the first corpus evaluation are significantly higher than those for the second corpus evaluation (t-Test:

$t = -3.574$, $df = 22.733$, $p < 0.001$), i.e., while participants in the first experiment aligned more to themselves, participants in the second experiment aligned more to their interlocutors. This noteworthy result indicates that the model actually reflects the participants alignment behaviour: Participants in the first experiment were focussed on *their* object names since they activated the corresponding lexical representation through explicit learning prior to the game. Participants in the second experiment on the contrary did not have such highly activated lexical representations and thus they aligned to their interlocutor more easily.

6 Discussion

The evaluation shows that SPUD *prime* and its underlying priming-based model of alignment are capable of simulating alignment phenomena found in psycholinguistic studies. Modelling lexical and syntactic alignment, as well as the effects of recency and frequency of use, it can account for a high degree of the lexical choices participants made in the two 'Jigsaw Map Game' tasks. A few points merit closer inspection.

First, it must be noted that the participants' behaviour (i.e., the behaviour producing the least number of mismatches) could in general be simulated not only by a single point or a compact cluster of points in parameter space, suggesting that the parameters are either not completely independent or that the evaluation method is too simple. Future empirical evaluations should have a wider scope and go beyond the generation of lexical items. SPUD *prime* could for instance be evaluated generating more sophisticated referring expressions. Yet, having several points in parameter space that achieve a certain behaviour is not problematic in an application context where the parameters can simply be set according to the theoretical model and the desired behaviour.

Second and interestingly, the parameters that lead to minimal numbers of mismatches in simulations can differ considerably between participants. Individual differences exist in verbal and non-verbal behaviour: Dale and Viethen [11] (this volume) report individual variation between referring expressions produced by human subjects describing a scene of geometrical objects and Bergmann and Kopp observe in their analysis of an extensive corpus of speech and gesture data that speakers differ significantly in the way they produce iconic gesture [2]. It can be expected that individual differences exist in speakers' alignment behaviour, too (cf. [12] for evidence). Here, data about participants' personalities should be collected so that a correlation with personality traits is possible.

Third, our model could generate a high number of the target nouns correctly, but failed on 10–20%. It should be noted, however, that it tries to give a purely mechanistic explanation of lexical and syntactic choice (in the spirit of Pickering and Garrod's interactive alignment model [19]) and that it, therefore, cannot explain alignment phenomena that are due to social factors (e.g., politeness, relationship, etc.), audience design or cases in which a speaker consciously decides whether to align or not (e.g., whether to use a word or its synonym). This is the main difference between our priming-based alignment model and the model of Janarthanam and Lemon [14] (this volume), which treats alignment from an audience design perspective (cf. Clark [10]). We think that a comprehensive model of alignment that accounts for all phenomena must unify both perspectives: low-level mechanistic alignment that is both rapid and broad in scope as well as more high-level strategic alignment that can account for audience design and social practices. How these two types of alignment could interact and influence each other is an open question.

7 Conclusion

In this paper, we introduced a priming-based model of alignment that focusses more on the psycholinguistic aspects of interactive alignment, and models recency and frequency of use effects — as proposed by Reitter [21] and Brennan and Clark [8] — as well as the difference between intrapersonal and interpersonal alignment [19, 18]. The presented model is fully parameterisable and can account for different empirical findings and 'personalities'. It has been implemented in the SPUD *prime* microplanner which activates linguistic rules by changing its knowledge base on-line and considers the activation values of those rules used in constructing the current utterance by using their mean activation value as an additional feature in its state evaluation function.

We evaluated our alignment model and its implementation in SPUD *prime* on two corpora of task-oriented dialogue collected in experimental setups especially designed for alignment research. The results of this evaluation show that our priming-based model of alignment is flexible enough to simulate the alignment behaviour of different human speakers (generating target nouns) in the experimental settings. Our model can reproduce human alignment behaviour to a high degree, but it remains to be investigated which influence each parameter exerts and how exactly the parameters vary across individual speakers.

Nevertheless, the development of the alignment-capable microplanner is only one step in the direction of an intuitive natural language human–computer interaction system. In order to reach this goal, the next step is to combine SPUD *prime* with a natural language understanding system, which should ideally work with the same linguistic representations so that the linguistic structures used by the interlocutor could be primed automatically. This work is underway.

Furthermore, user studies should be carried out in order to evaluate SPUD *prime* in interactive scenarios. Branigan et al. [6] found that human–computer alignment was even stronger than human–human alignment. But how would the alignment behaviour of human interlocutors change if the computer they are speaking to also aligns to them? Further, would integration of an alignment-capable dialogue system into a computer interface make the interaction more natural? And would an embodied conversational agent appear more resonant and more sociable [16], if it aligned to users during conversation? The work presented here provides a starting point for the investigation of these questions.

Acknowledgements. This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence in 'Cognitive Interaction Technology' (CITEC) as well as in the Collaborative Research Center 673 'Alignment in Communication'. We thank Petra Weiß for making the 'Jigsaw Map Game' corpora available and Mariët Theune for some very helpful comments on the draft of this chapter.

References

1. Bateman, J.A.: A social-semiotic view of interactive alignment and its computational instantiation: A brief position statement and proposal. In: Fischer, K. (ed.) How People Talk to Computers, Robots and Other Artificial Communication Partners. pp. 157–170. SFB/TR 8 Report No. 010-09/2006, Bremen, Germany (2006)
2. Bergmann, K., Kopp, S.: GNetIc – Using bayesian decision networks for iconic gesture generation. In: Proceedings of the 9th International Conference on Intelligent Virtual Agents. pp. 76–89. Amsterdam, The Netherlands (2009)
3. Bock, J.K.: Syntactic persistence in language production. Cognitive Psychology 18, 355–387 (1986)
4. Bock, J.K., Griffin, Z.M.: The persistence of structural priming: Transient activation or implicit learning? Journal of Experimental Psychology: General 129, 177–192 (2000)
5. Branigan, H.P., Pickering, M.J., Cleland, A.A.: Syntactic priming in written production: Evidence for rapid decay. Psychonomic Bulletin & Review 6, 635–640 (1999)
6. Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F.: Linguistic alignment between people and computers. Journal of Pragmatics (in press)
7. Brennan, S.E.: Conversation with and through computers. User Modeling and User-Adapted Interaction 1, 67–86 (1991)
8. Brennan, S.E., Clark, H.H.: Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition 22, 1482–1493 (1996)
9. Brockmann, C., Isard, A., Oberlander, J., White, M.: Modelling alignment for affective dialogue. In: Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling. Edinburgh, UK (2005)
10. Clark, H.H.: Using Language. Cambridge University Press, Cambridge, UK (1996)
11. Dale, R., Viethen, J.: Attribute-centric referring expression generation. In: Krahmer, E., Theune, M. (eds.) Empirical Methods in Natural Language Generation, LNCS, vol. 5980. Springer, Berlin / Heidelberg (2010)
12. Gill, A.J., Harrison, A.J., Oberlander, J.: Interpersonality: Individual differences and interpersonal priming. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society. pp. 464–469. Chicago, IL (2004)
13. Isard, A., Brockmann, C., Oberlander, J.: Individuality and alignment in generated dialogues. In: Proceedings of the 4th International Natural Language Generation Conference. pp. 25–32. Sydney, Australia (2006)
14. Janarthanam, S., Lemon, O.: Learning adaptive referring expression generation policies for spoken dialogue systems. In: Krahmer, E., Theune, M. (eds.) Empirical Methods in Natural Language Generation, LNCS, vol. 5980. Springer, Berlin / Heidelberg (2010)

15. de Jong, M., Theune, M., Hofs, D.: Politeness and alignment in dialogues with a virtual guide. In: Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems. pp. 207–214. Estoril, Portugal (2008)
16. Kopp, S.: Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication (accepted manuscript)
17. Levelt, W.J.M., Kelter, S.: Surface form and memory in question answering. Cognitive Psychology 14(1), 78–106 (1982)
18. Pickering, M.J., Branigan, H.P., McLean, J.F.: Dialogue structure and the activation of syntactic information. In: Proceedings of the 9th Annual Conference on Architectures and Mechanisms for Language Processing. p. 126. Glasgow, UK (2003)
19. Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences 27(2), 169–226 (2004)
20. Purver, M., Cann, R., Kempson, R.: Grammars as parsers: Meeting the dialogue challenge. Research on Language and Computation 4, 289–326 (2006)
21. Reitter, D.: Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora. Ph.D. thesis, University of Edinburgh (2008)
22. Stone, M.: Lexicalized grammar 101. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. pp. 77–84. Philadelphia, PA (2002)
23. Stone, M., Doran, C., Webber, B., Bleam, T., Palmer, M.: Microplanning with communicative intentions: The SPUD system. Computational Intelligence 19, 311–381 (2003)
24. Weiß, P., Pfeiffer, T., Schaffranietz, G., Rickheit, G.: Coordination in dialog: Alignment of object naming in the Jigsaw Map Game. In: Proceedings of the 8th Annual Conference of the Cognitive Science Society of Germany. pp. 4–20. Saarbrücken, Germany (2008)
25. Weiß, P., Pustylnikov, O., Mehler, A., Hellmann, S.M.: Patterns of alignment in dialogue: Conversational partners do not always stay aligned on common object names. In: Proceedings of the Conference on Embodied and Situated Language Processing. p. 16. Rotterdam, The Netherlands (2009)