UNIVERSITY OF CALIFORNIA

SANTA CRUZ

LINGUISTIC ALIGNMENT IN NATURAL LANGUAGE GENERATION

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

GABRIELLE M. HALBERG

September 2013

The Thesis of Gabrielle M. Halberg
is approved:

_____

Professor Marilyn Walker, Chair

_____

Professor Sri Kurniawan

_____

Professor Arnav Jhala

_____

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Linguistic Alignment in Natural Language Generation

by

Gabrielle Halberg

Linguistic alignment in dialogue refers to the tendency of conversational partners to mutually align their language and speech patterns. This behavior is considered to be a natural and productive aspect of verbal communication as it contributes significantly to the overall communicative success of a conversation. Current methods of natural language generation in deployed dialogue systems do not make use of this feature. In this work we develop and evaluate PERSONAGE-*primed*, a natural language generation system capable of performing dynamic linguistic alignment in the task domain of walking directions. This system builds upon an existing language generation method by extending functionality and modifying input parameters. An evaluation of the system's output revealed that most generated utterances were perceived to sound natural within existing human dialogues but that certain linguistic features seemed to contribute negatively to the perceived naturalness of the utterance. This work provides a strong foundation for further investigation of which features are relevant for implementing productive linguistic alignment. It also provides insight as to how methods of natural language generation might improve some forms of human-computer interactions.

Supported in part by NSF CISE-IIS Grant #1044693.

Special thanks to Professor Marilyn A. Walker for her direction, encouragement, and valuable input.

1 Introduction

The presence and widespread use of spoken dialogue systems in our digital culture is increasing. Recent improvements in speech recognition technologies has made them a more viable solution for successful human-computer interactions. However, in spite of these improvements many existing dialogue systems still seem awkward and challenging to use when they fall short of our natural, conversational expectations. By implementing in dialogue systems patterns observed to occur in natural conversations, the interactions are likely to be more intuitive, user-friendly and subsequently less cognitively demanding. In this work we make progress towards this goal by developing a language generation system capable of performing linguistic alignment.

Linguistic alignment refers to the tendency of speakers to mutually align their linguistic patterns over the course of a dialogue. For example, a study from 1982 by Levelt and Kelter [14] provides a simple demonstration of this behavior. In their study they telephoned shops and inquired about the shop hours in one of two ways: If the caller asked, "At what time does your shop close?" the clerk would generally respond with, "At 6 o'clock", using the same preposition used by the caller; If the caller asked instead, "What time does your shop close?" the clerk would generally respond with a simple, "6 o'clock", this time omitting the preposition. In these examples the caller's linguistic choices influence those of the clerk.

It is precisely these subtleties of dialogue that we seek to better understand in order to improve existing methods of designing language interactions. Towards that goal, this paper presents a method of alignment-capable natural language generation in the domain of pedestrian walking directions. While research has shown that linguistic alignment is observed to occur on various linguistic levels including phonological and prosodic forms [11], this work focuses specifically on generating lexical, syntactic and stylistic alignment. The objective of this work is to present and evaluate our method of alignment-capable language generation and create a foundation for future research on the effect of linguistic alignment in dialogue. Section 2 frames this research in the context of other work on linguistic alignment and

natural language generation. Section 3 provides a detailed description of the system and method of implementation. Section 4 describes the evaluation experiment and section 5 presents the results. Section 6 offers a discussion and section 7 provides some concluding thoughts.

2 Background and Related Work

Pickering and Garrod [11] present a comprehensive theory of the underlying mechanisms and motivations for alignment, which they call the interactive alignment model (IAM). Their model explains that there is an unconscious mutual influence between language comprehension and production, which is both a reason for and an effect of the presence of linguistic alignment in conversation. They assert that successful dialogue is dependent on this alignment of representations between dialogue partners. In support of this claim, several studies of task-oriented dialogues (i.e. in which the conversants must communicate to solve a problem together) have revealed a correlation between the presence of linguistic alignment and the overall task success [20] [25] [22] [18]. Garrod and Anderson [10] addressed conceptual and semantic coordination through a collaborative maze game. They found that users converged to use the same syntactic forms of expressing location, which supports Pickering and Garrod's claims about alignment of representations on various levels.

Since Levelt and Kelter's early question-answering study, numerous additional studies have shown evidence of alignment in both human-human interactions [2] [1] and human-computer interactions [23] [19]. Additional work with *Let's Go!* [21], a telephone-based spoken dialogue system for information on bus routes, showed that users do align to the system's lexical choices and concept forms. This work and various other research on measuring alignment provides a target model for how to appropriately emulate the behavior within a dialogue system.

In addition to this work on *measuring* linguistic alignment, there exists a variety of prior work on *generating* linguistic alignment. In 2010 Mairesse and Walker first presented PERSONAGE [17], a highly parameterizable generator with parameters to support adaptation to a user's linguistic style. Jong et al. [7] presents an approach

that focuses on affective language use for aligning specifically to user's politeness and formality. Brockman et al. [3] illustrates a model in which alignment is simulated using word sequences alone. An extension of this work in Isard et al. [13] simulates both individuality and alignment in dialogue between pairs of agents with the CrAg-2 system; This system uses an over-generation and ranking approach that yields interesting results, but the underlying method has no explicit parameter control and the output has yet to be evaluated.

Most relevant is the alignment-capable microplanner SPUD *prime* presented by Buschmeier et al [4]. SPUD *prime* is a computational model for language generation in dialogue that focuses heavily on relevant psycholinguistic and cognitive aspects of the interactive alignment model. Their system is driven by a method of activating relevant rules in a detailed contextual model according to user behavior during a dialogue. Although the underlying system seems to be capable of producing both syntactic and lexical alignment, it was evaluated only for accurate representation of lexical alignment in a corpus of dialogues from a controlled experiment.

3 Method

In this section we describe our alignment-capable natural language generation system Personage-*primed*, which is an extension of the parameterizable language generator Personage [16]. The Personage system was initially designed to generate utterances to express targeted personalities based on particular linguistic features in the domain of restaurant recommendations. Personage is capable of producing a wider range of linguistic variation than traditional template-based language generation systems because it dynamically modifies high level representations of the utterances and implements external lexical resources including Verbocean [5] and Wordnet [8]. In Personage-*primed* however, the lexical values are derived instead directly from a set of prime values, which represent the content and linguistic information from a dialogue to which the generated utterance is a response - that is, the output sentence is generated to align with the given dialogue. In future work Personage-*primed* should be extended to include a mixed method that draws from

both the available lexical and aggregation resources as well as the prime values in order to emulate how both existing language models and contextual linguistic information can influence language production.

The basic architecture of PERSONAGE-*primed* is shown in Figure 1, which is explained in further detail in the following sections. The system output is a complete utterance that presents the communicative goal in alignment with the lexical, structural and stylistic information from the prime values.

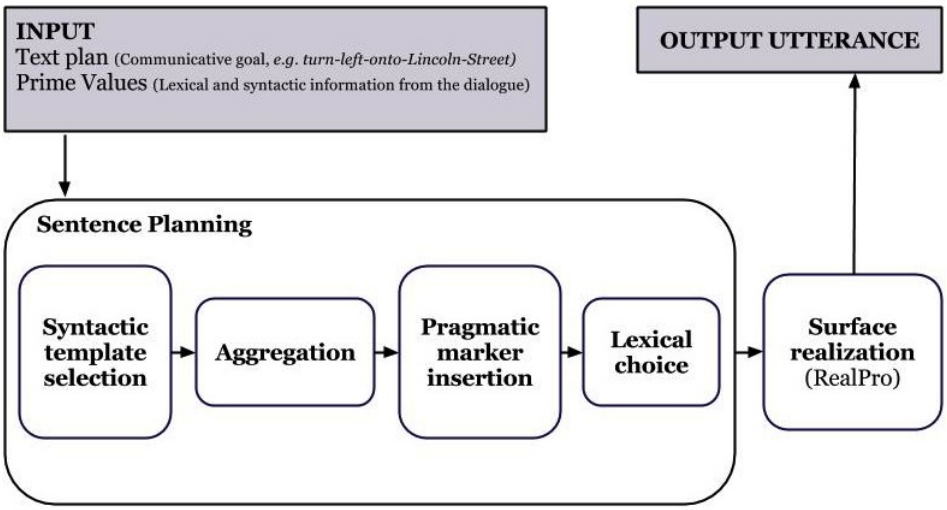


Figure 1: Basic architecture of PERSONAGE-*primed*

3.1 Input

The input to PERSONAGE-*primed* consists of a text plan and a set of alignment target values referred to as the prime values. The prime values contain lexical and syntactic information from the dialogue to which the generated utterance will be aligned. The text plan is a high level structure representing the communicative goal of the desired output utterance. Each text plan contains either a single instruction or a compound instruction. A compound instruction consists of two clauses (an instruction or statement) joined by a temporal relation, such as *after*, *until* or *once*. An example text plan tree for a compound instruction is shown in Figure 2.

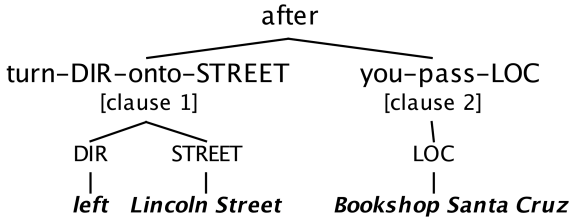


Figure 2: Example text plan tree. Bold lexemes in italic are variables that are instantiated at generation time.

PERSONAGE-*primed* currently supports 13 unique instructions and statements for the walking directions domain, but could easily be extended to include more. Table 1 contains a complete list of the supported instructions and statements.

Instruction/Statement	Example Utterance
confirm (yes)	*That's correct.*
turn-DIR	*Make a right turn.*
turn-DIR-onto-STREET	*At Cedar Street, make a right.*
continue-on-STREET	*Keep going straight down Cedar Street.*
continue-on-STREET-for-NUM-blocks	*You're going to follow Cedar Street for three more blocks.*
go-along-STREET	*Head down Cedar Street.*
go-NUM-blocks-on-STREET	*Walk five blocks along Cedar Street.*
go-to-LOC-from-LOC	*... you will walk from the bookstore to the coffeeshop.*
go-to-LOC-from-STREET-and-STREET	*From the corner of Cedar Street and Elm, walk towards Lulu's Coffeeshop.*
go-back-to-LOC	*Go back to the bookstore.*
you-pass-LOC	*After you pass the bookstore...*
LOC-is-on-the-DIR	*The bookshop is on the left-hand side.*
arrive-at-LOC	*When you get to Lincoln Street...*

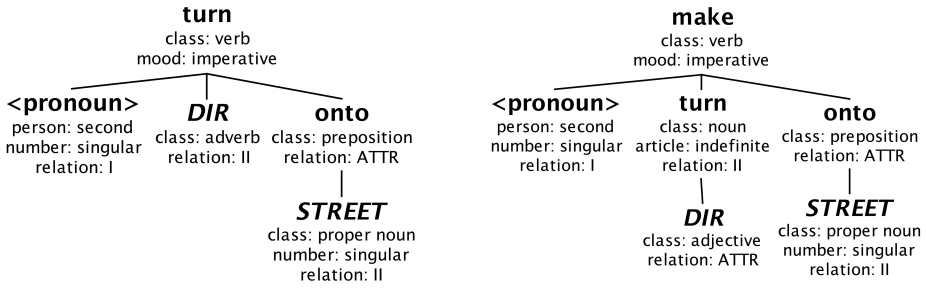
Table 1: Instructions and statements supported in PERSONAGE-*primed*

3.2 Syntactic Template Selection

While the text plan contains all the information regarding what will be communicated, the sentence planning pipeline controls how that information is conveyed.

The syntactic template selection is the first phase of sentence planning that selects the most appropriate syntactic form for the instruction(s) in the text plan.

In order to properly represent and manipulate the syntactic form of a sentence there must be an associated data structure. PERSONAGE-*primed* implements the same syntactic dependency tree representation for utterances as used in PERSONAGE [17], referred to as a Deep Syntactic Structure (DSyntS). The DSyntS specifies the relationship between the different components of a sentence. Two example DSyntSs for the instruction *turn-DIR-onto-STREET* are shown in Figure 3. In these examples, the verb is the root of the tree. All other components or phrases of the sentence are children of the verb and are assigned relation values. The relation I indicates that the component is the subject of the parent. The relation II indicates that the component is the direct object of the parent. The relation III indicates that the component is the indirect object of the parent. The relation ATTR indicates that the component is a modifier (such as an adjective or a prepositional phrase) of the parent. The DSyntS data structure is an important aspect of the PERSONAGE-*primed* system as it allows for appropriate manipulation of the utterance further down the sentence planning pipeline.



(a) Turn DIR onto STREET (b) Make a DIR turn onto STREET

Figure 3: Two example DSyntSs for the instruction *turn-DIR-onto-STREET*. The lexemes are in bold, and the attributes below indicate non-default values in the RealPro Realizer. Bold lexemes in italic (e.g ***DIR*** and ***STREET***) are variables that are instantiated at generation time.

The DSyntSs are stored in a handcrafted generation dictionary. While creating this generation dictionary is the most labor-intensive aspect of the PERSONAGE-*primed* system, the process could be simplified by incorporating a method of au-

tomatically populating the dictionary, such as the unsupervised learning approach described in Higashinaka et al. [12].

Each instruction and statement has an associated DSyntS List, which is a collection of semantically equivalent DSyntS with different structures. This relationship is illustrated in Figure 4 in which each instruction/statement of the given text plan points to a DSyntS List which is a collection of DSyntSs. Because some communicative goals have more potential variation than others, some DSyntS Lists are larger than others.

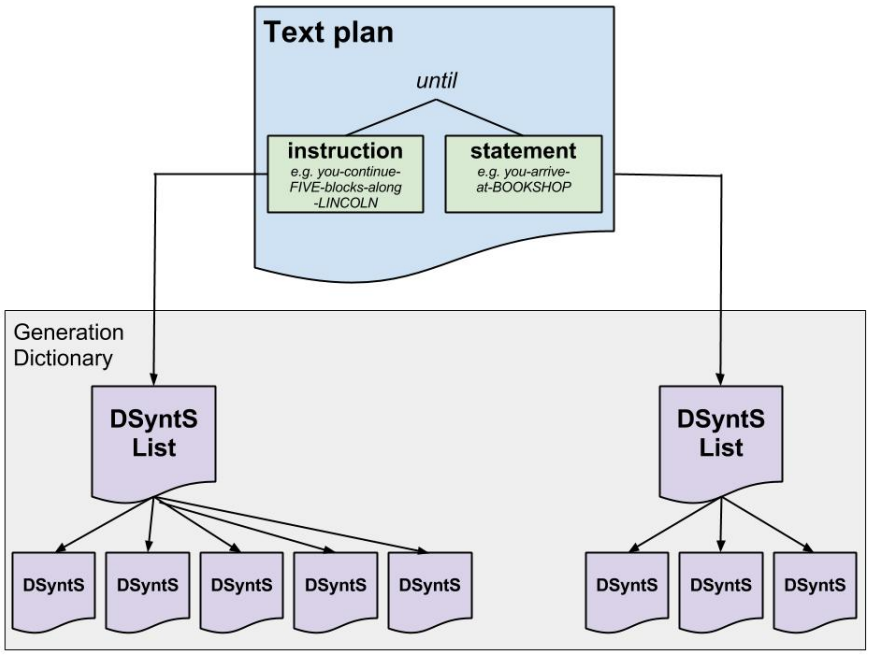


Figure 4: Illustration of the relationship between the content of a text plan and the associated DSyntS List.

During syntactic template selection, for each instruction in the text plan PERSONAGE-*primed* finds the associated DSyntS List and selects the DSyntS that best matches the lexical and syntactic information in the prime values. This best match is determined according to a set of tagged features on each DSyntS - the DSyntS with the highest number of features matching the prime values is designated as the best match. If no best match is found, the default DSyntS is assiged to the instruction. The DSyntS variables NUM, LOC, STREET, and DIR are filled at generation time from the content given in the text plan.

Each DSyntS in the DSyntS List is tagged with features that distinguish it from the others in the list. Many features are distinguishing lexical values, but those which share too many lexical features are tagged for their unique syntactic structure. Instructions including multiple prepositions may have two or more possible orderings. For example, the instruction "Head over to the bookshop from Cedar Street and Elm" may also be realized as "From Cedar Street and Elm, head over to the bookshop." The DSyntS for the first form is tagged with **VP-PP-PP** to indicate that the verb phrase is followed by two prepositional phrases, and the DSyntS for the second form is tagged with **PP-VP-PP** to indicate that one prepositional phrase precedes the verb phrase. It is important to distinguish syntactic forms like this to account for syntactic alignment in dialogue. If a navigation dialogue included the question, "From here where should I go to next?" a response with syntactic alignment would be phrased in a similar way, such as "From where you are, walk to Pacific Avenue and then make a left."

3.3 Aggregation

PERSONAGE-*primed* further extends the PERSONAGE generation system by introducing new clause combining operations for each of the new temporal relations. These operations are executed during the aggregation phase. For compound instructions that contain a temporal relation (such as $after$ or *once*), the aggregation component integrates each DSyntS into a larger syntactic structure. For most temporal relations, the clauses can be joined in two ways; the relation can appear at the beginning of the sentence followed by the two clauses it relates, as in "**After** you pass the bookshop, turn left onto Cedar Street"; or the relation can occur in between the two clauses as in "Turn left onto Cedar Street **after** you pass the bookshop". The temporal relation "and then" is one exception to this pattern as it never appears sentence initially. The clause combining operations control how this aggregation is carried out.

3.4 Pragmatic Marker Insertion

Following the aggregation operation is pragmatic marker insertion. Pragmatic markers, or discourse markers, are elements of spontaneous speech that do not necessarily contribute to the semantic content of a discourse but instead serve to smooth a conversation in various ways. Some common examples include "so", "okay, "like", "umm", "you know" and "yeah (not in response to a yes / no question). Research on spontaneous speech has shown that discourse markers not only make a conversation sound more natural but can also serve to highlight or qualify content, help listener's follow a speaker's train of thought, and create a meaningful transition from one utterance to the next [9] [6]. If discourse markers were meaningless they would certainly be less prevalent in spoken dialogue.

Discourse markers are especially prelevant in task-oriented dialogue. In PERSONAGE-*primed* the pragmatic marker insertion phase will insert up to three[1] of the pragmatic markers found in the prime values.

Each pragmatic marker has an associated a set of possible insertion points - a set of rules or patterns for use. For example, "so" is generally sentence initial, while "like" and "you know" can occur between phrases. During this phase of the generation process, a pragmatic marker is inserted only if one of the insertion points associated with the marker is present in the DSyntS. In addition to the pragmatic markers, PERSONAGE-*primed* extends this phase of generation by also inserting adverbial modifiers such as "next" and "now" if they are present in the prime values.

3.5 Lexical Choice

The lexical choice operation encompasses several finer-grained operations, including modifier insertion, synonym selecton, referring expression selection, tense transformation and modal insertion. Lexical choice is the final step of sentence planning prior to surface realization.

[1] While use of pragmatic markers varies according to individual personalities, three was chosen to be a maximum value as it reflected an approximation of average use.

3.5.1 Synonym Selection

The synonym selection operation checks every verb and preposition in the current utterance and if there exists a synonym in the prime values, the prime synonym replaces the existing verb or preposition. The system does not currently align to nouns because most nouns within the walking directions domain are referring expressions, such as "downton", "Pacific Avenue", etc. Alignment to referring expressions is handled with a separate operation. In addition, many common nouns in the directions domain do not have appropriate synonyms, such as directions like "right" and "left". However, if a system was equipped with detailed knowledge of an area, including common objects and landmarks such as sculptures, murals, or details about storefronts, alignment to common nouns and descriptions in particular would be more relevant.

3.5.2 Referring Expression Selection

The referring expression selection is very similar to the synonym selection operation. Its main function is to check every proper noun within the current utterance for a semantic match in the prime values. This operation requires an existing database of referring expressions and their possible variations. For this work we manually created a map from each referring expression to its list of variations. For example, the destination named "Bookshop Santa Cruz" is an entry in the referring expression map with the corresponding list of alternative referring expressions {"bookshop", "the bookshop", "Santa Cruz bookshop"}.

In addition to accounting for variation of referring expressions, this operation also accounts for a referring expression form that is commonly found in navigation dialogues - referencing street names without the street suffix. If one participant in a dialogue refers to a street as "Pacific" instead of "Pacific Avenue", it is common for the other participant to do so as well. This step of the referring expression operation checks the prime values for any single instance of this shortened form and modifies all instances of street names in the current utterance to align with this stylistic choice. This addition is based on research that reported measurable alignment with

time forms, e.g. referencing a time as “six”, “six a.m.”, ‘six o’clock‘, etc [23].

3.5.3 Tense transformation and modal insertion

While there is no specific research regarding whether or not alignment of tense and modals is common in dialogues, it is nonetheless an instance of lexical alignment and so we designed the system to have this capability. These operations are fairly simple; If there exists an explicit use of a particular tense or a modal in the prime values, the current utterance is modified to align. The simple future tense (discussed in further detail below) is the most complicated instance of this operation as it requires a structural transformation of the utterance; that is, the current verb phrase gets embedded as an infinitive phrase within a new verb phrase.

The most common tenses used for giving directions in the navigation domain are present, future, and simple future. While followers do often use past tense to confirm the completion of an action, such as “Okay I went three blocks”, it is less common for directors to use it and so we did not include a specific transformation to generate it. The present tense is most often used with imperative commands that lack an explicit subject; A director will often say “Go five blocks along River Street”. The future tense often arises in response to questions such as “How long will I be on Lincoln (Street)?” The simple future tense is a common alternative construction for expressing future events; instead of “You will turn left on Pacific (Avenue)” it’s “You are going to turn left on Pacific (Avenue).” The use of simple future is more common for directors than for followers.

The modals “should”, “can” and “might” are commonly found in navigation dialogues. Followers will express uncertainty with questions such as “Should I stay on Pacific Avenue?” or seek confirmation for alternative routes with questions like “Can I take Cooper all the way to Pacific?”. The corresponding director responses sometimes align with this lexical addition with confirming responses such as “Yes, you should stay on Pacific Avenue for three more blocks” or “Yes, you can take Cooper all the way.” The tense transformation and modal insertion operations are intended to reflect these observed patterns.

4 Experiments

To evaluate the naturalness of the utterances generated by PERSONAGE-*primed* we designed a simple perceptual experiment in the form of a short survey. Responses to the survey were collected via Amazon Mechanical Turk. The survey included ten questions. For each question, participants were presented with an excerpt from a dialogue in which a director (D) is instructing a follower (F) how to navigate to a destination on foot. The dialogue excerpts were taken from the Art Walk corpus [15] and were slightly modified to isolate certain priming values. Following the excerpt, participants were presented with three options for what the director could say next; one option was a default (non-aligned) utterance generated from PERSONAGE-*primed* without any priming values; a second option was the aligned utterance generated from PERSONAGE-*primed* with the priming values and the same textplan as used for the default utterance; the third option was an utterance - with the same communicative goal as those generated from the system - randomly selected from a human dialogue in the ArtWalk corpus [15]. For each option, the participants were required to rate how naturally each utterance followed from the provided dialogue and context, on a scale from 1 (very awkward) to 5 (very natural). An example question is shown in Figure 5. The questions and options were randomized between participants. There were 61 participants in total.

Figure 5: An example question from the experiment

5 Results

This experimental evaluation of PERSONAGE-*primed* revealed that most generated utterances were perceived to sound natural within existing human dialogues. Table 2 shows that 9 out of the 10 utterances presented in the survey were judged on average to be moderately natural with a mean of 3.50 out of 5. Table 3 summarizes the naturalness scores for the generated, non-aligned utterances, which had the highest average rating with a mean of 3.90 out of 5. The random human utterances shown in Table 4 had the lowest naturalness score with a mean of 3.38 out of 5. While these results are somewhat unexpected, the generated aligned utterances *were* rated at least as natural as the human utterances - on average they were even slightly better. This suggests that the generated aligned utterances do sound more natural than a random instruction given out of context. However, further research is required to understand the underlying factors for these mixed results.

	Aligned Utterance	**Mean**	**Median**	**Std. Dev**
1	Okay, you should make a left.	3.57	4	1.15
2	Okay, next, keep going on Pacific until you get to Walnut.	3.66	4	0.98
3	Yeah, at Cedar hang a right.	3.61	4	1.23
4	Yeah, okay, go like, towards Pacific.	2.98	3	1.09
5	Okay, so turn like, right onto Lincoln Street.	3.20	3	0.98
6	Okay, after you go past Cathcart, turn left on Lincoln.	3.66	4	0.98
7	Okay, so you will continue on Lincoln for three blocks.	3.57	4	1.01
8	Okay, it will be on the left side.	3.56	4	1.01
9	Yeah, so you are going to walk two more blocks along Front and then it's towards the left.	3.52	4	1.04
10	From Cedar and Elm walk towards downtown.	3.64	4	0.91
	Summary Statistics			
	Mean			**3.50**
	Median			**4**
	Standard Deviation			**1.05**

Table 2: Summary of naturalness scores for the generated aligned utterances

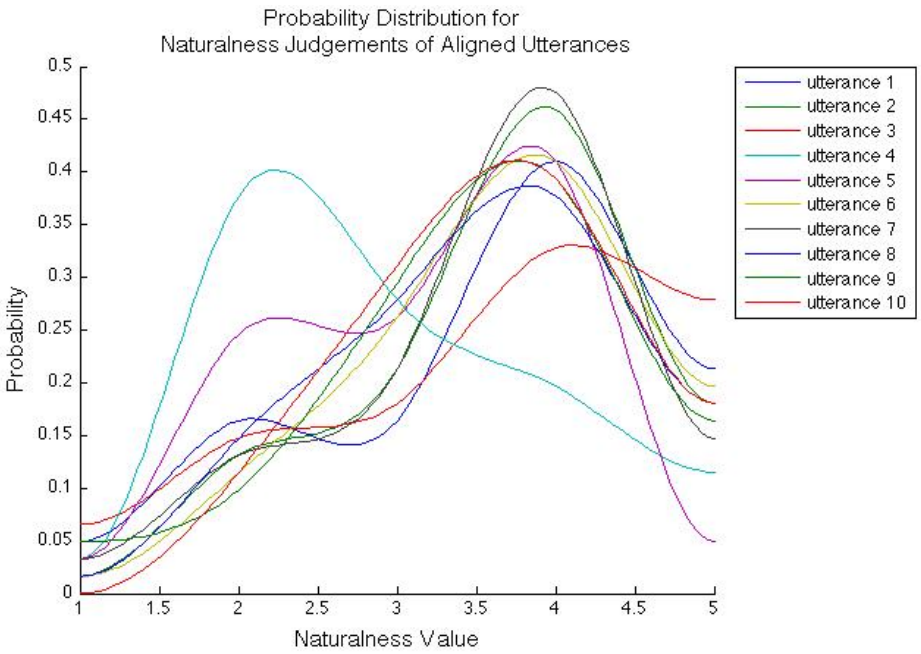


Figure 6: Probability distributions for all aligned utterances

A closer look at the specific distribution of naturalness judgements serves to inform future work. The results suggest that certain linguistic features may contribute negatively to perceived naturalness. The only two aligned utterances that received an average score below 3.5 - utterances 4 and 5 - both contained a lexical feature unique to the set of utterances from the experiment: the use of the pragmatic marker, "like". Utterance 4 received the lowest average score of all the aligned utterances. The pragmatic marker in utterance 4 was inserted immediately preceding a preposition, which is not commonly used. In addition, it was used in an identical context as in the immediately previous dialogue turn, which may have created a kind of mocking effect. In contrast, in utterance 5 the pragmatic marker immediately precedes an adverb, which is a more common pattern and it is used in a different context than the previous dialogue turn. This potential preference is illustrated in the probability distributions shown in Figure 6.

Utterance 4 has a single large cluster around the 2.3 rating. The distribution is shown in greater detail in Figure 7. In contrast, utterance 5 has two clusters - one centered around the 2.2 rating and another centered around the 3.8 rating, shown in greater detail in Figure 8. This bimodal distribution suggests that while

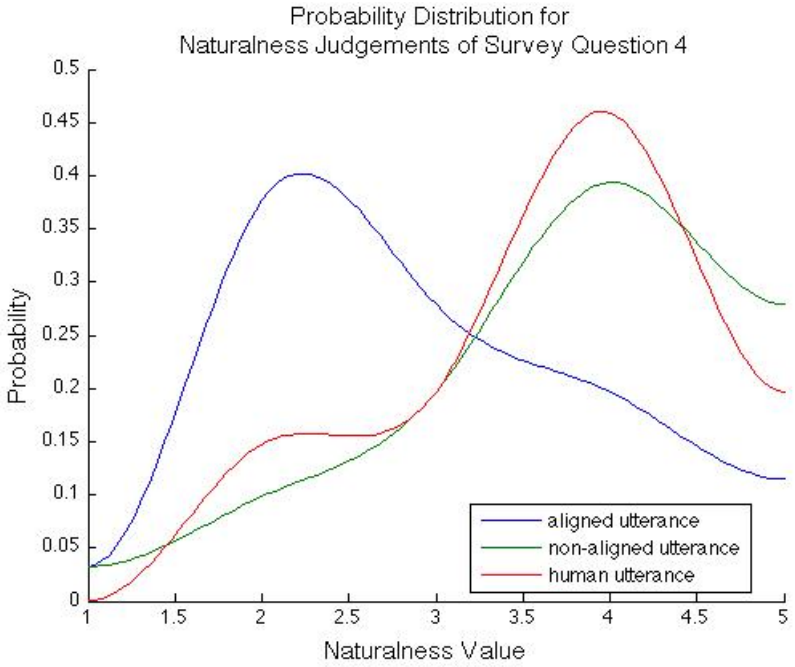


Figure 7: Probability distributions for survey question 4

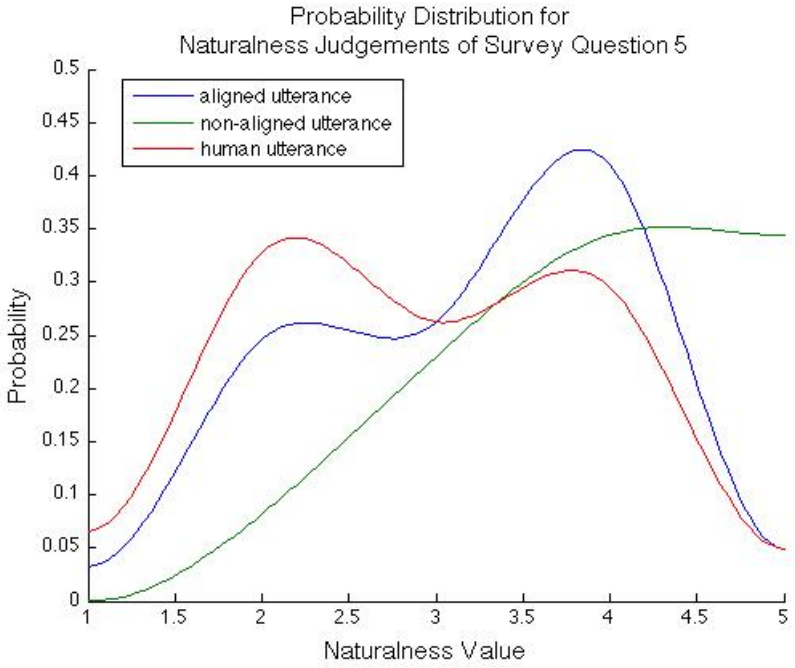


Figure 8: Probability distributions for survey question 5

some participants perceived the utterance to be unnatural, a larger group found the utterance to be rather natural. However, the various differences between the two sentences and their prime values suggest that there may be more than one influencing

factor and that further research is required to isolate their effects.

	Non-aligned Utterance	Mean	Median	Std. Dev
1	Turn left.	4.10	4	0.96
2	Continue on Pacific Avenue until you arrive at Walnut Avenue.	4.02	4	1.02
3	Yes, turn right onto Cedar Street.	3.93	4	0.85
4	Yes, go to Pacific Avenue.	3.79	4	1.07
5	Turn right onto Lincoln Street.	3.95	4	0.96
6	After you pass Cathcart Street, turn left onto Lincoln Street.	3.89	4	0.93
7	Continue on Lincoln Street for three blocks.	3.89	4	0.97
8	It's on the left.	4.11	4	1.03
9	Go two more blocks along Front Street and then it's on the left.	3.70	4	1.01
10	Go to downtown from Cedar Street and Elm Street.	3.61	4	1.13
	Summary Statistics			
	Mean			**3.90**
	Median			**4**
	Standard Deviation			**1.00**

Table 3: Summary of naturalness scores for the non-aligned (default), generated utterances

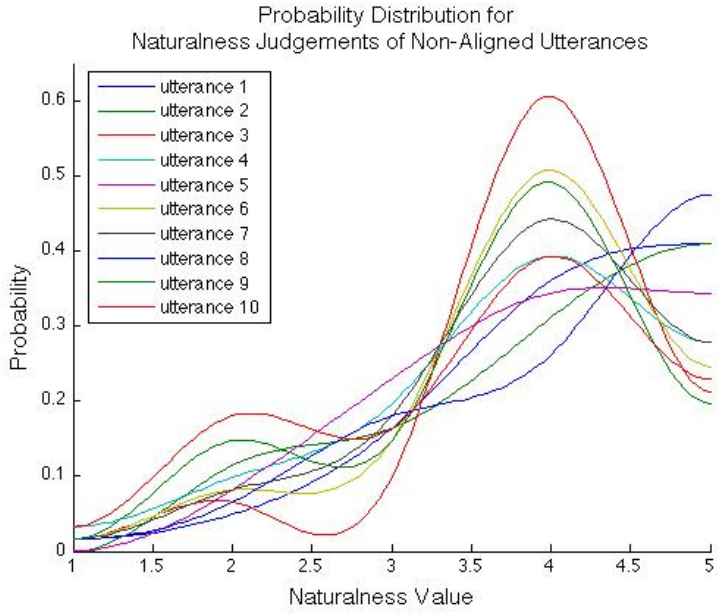


Figure 9: Probability distributions for all non-aligned utterances

The probability distributions for the non-aligned utterances shown in Figure

9 show less variation across judgements (although there are still some very small clusters on the low end of the naturalness spectrum). This is perhaps a result of the straight-forward nature of these sentences. With less stylistic variation there is less variation among judgements. In contrast, the distributions for the human utterances shown in Figure 10 show consistent bimodal distributions, which is potentially a result of the highly stylized and colloquial nature of those sentences.

	Human Utterance	**Mean**	**Median**	**Std. Dev**
1	Go left.	4.05	4	1.04
2	Okay and then you're gonna go down Pacific Avenue until you hit Walnut.	3.31	3	0.94
3	Yeah, so you wanna turn right onto Cedar.	3.30	3	0.95
4	Yeah, so continue going to Pacific.	3.70	4	0.95
5	So like once you get to Lincoln Street turn right onto it.	2.93	3	1.05
6	Lincoln is the next left uh, that you can make after Cathcart, I think.	2.85	3	1.00
7	And then keep going, keep going down Lincoln for three blocks.	3.23	3	1.06
8	It's on the left hand from the way you are going.	3.49	4	1.18
9	Oh great, you're closer, couple more blocks, and it's on the left side,	3.69	4	1.09
10	Okay, so you're heading to downtown from Cedar and Elm.	3.20	3	1.08
	Summary Statistics			
	Mean			**3.38**
	Median			**4**
	Standard Deviation			**1.09**

Table 4: Summary of naturalness scores for the human utterances

For each survery question we evaluated the significance of the differences between samples using a Kruskal-Wallis test. The Kruskal-Wallis significance test evaluates the null hypothesis that the differences are due to random sampling. A small p-value rejects the null hypothesis and suggests instead that at least one sample median is significantly different from the others. We used a significance level of 0.05. The results of the Kruskal-Wallis test for each survery question are shown in Table 5.

In order to determine which pairs within each sample were significantly different and which were not, we used a multiple comparison test between the three condi-

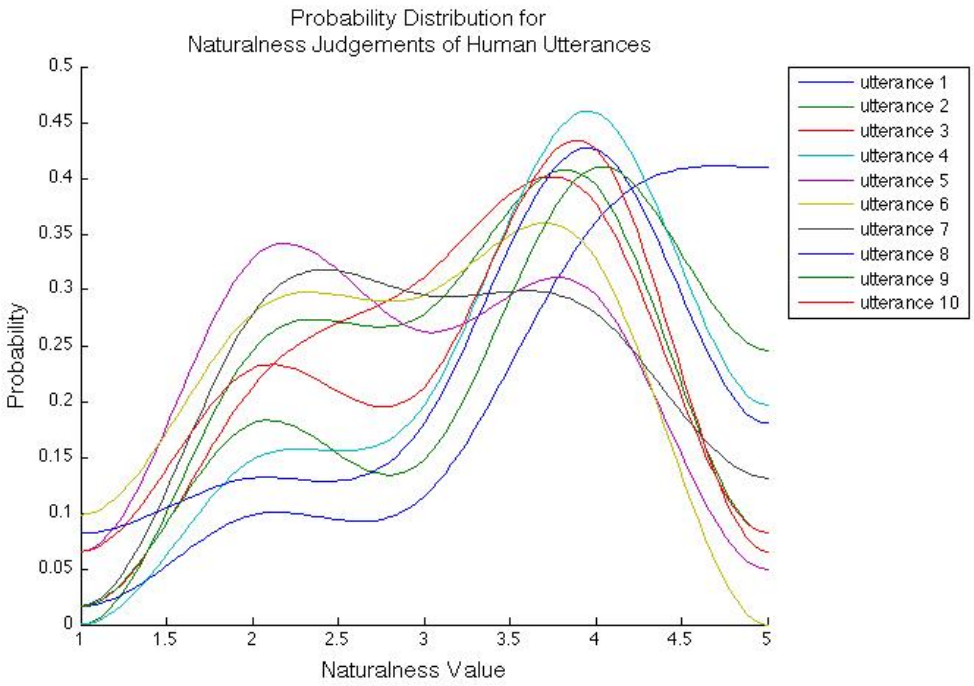


Figure 10: Probability distributions for all human utterances

Question	***p*-value**	**Aligned**	**Non-aligned**	**Aligned**	**Human**
1	**0.0127**	3.57	4.10	3.57	4.05
2	**0.0004**	3.66	4.02	3.66	3.31
3	**0.0013**	3.61	3.93	3.61	3.30
4	$\mathbf{4.245 \cdot 10^{-5}}$	**2.98**	**3.79**	**2.98**	**3.70**
5	$\mathbf{7.208 \cdot 10^{-7}}$	**3.20**	**3.95**	3.20	2.93
6	$\mathbf{1.227 \cdot 10^{-7}}$	3.66	3.89	**3.66**	**2.85**
7	**0.0016**	3.57	3.89	3.57	3.23
8	**0.0014**	**3.56**	**4.11**	3.56	3.49
9	0.4899	3.52	3.70	3.52	3.69
10	0.0540	3.64	3.61	3.64	3.20

Table 5: Results of a Kruskal-Wallis test and follow up multiple comparison test. Values shown in bold are those found to be significant at the 0.05 level.

tions presented for each question; aligned, non-aligned, and human-generated. A difference between two conditions is considered significant at the 0.05 level if the confidence interval for the true difference of the means does not contain 0.0 and the intervals of the two means are disjoint; two means are not significantly different if their intervals do overlap. The multiple comparison test revealed that there were only significant differences between naturalness scores across conditions in survey questions 4, 5, 6, and 8. Table 5 shows a summary of this analysis with respect to

the two conditions of interest; aligned versus non-aligned and aligned versus human. Means highlighted in bold are those that were found to be significantly different according to the multiple comparison test at the 0.05 level.

For question 4 (shown in Figure 11), there was a significant difference between the mean naturalness scores for the aligned and the non-aligned utterance as well as between the aligned and the human utterance. The aligned utterance received a mean score of 2.98 while the non-aligned and the human utterances received mean scores of 3.79 and 3.7, respectively. A qualitative analysis of the question and the three presented utterances reveals that the repetitive use of the pragmatic marker "like" is unique amongst the three conditions; it is the only one of the three options to contain the exact construction, "like, towards Pacific", that is used in the immediately preceding dialogue turn by the follower in the dialogue, which may create a mocking effect that is not perceived to be natural. There are two other features in the aligned utterance that are unique amongst the three options - the pragmatic marker "okay" and the preposition "towards". However, because these features do not appear to contribute negatively to the naturalness scores of many other utterances in the survey, it is reasonable to assume that they are not likely to have contributed significantly to the low naturalness score of this aligned utterance.

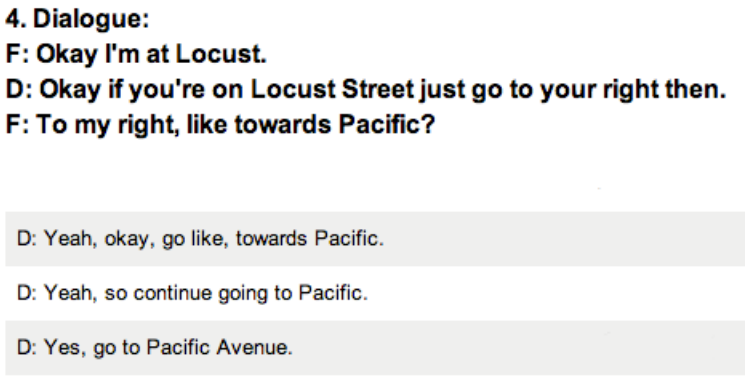


Figure 11: Survey question 4

Question 5 is shown in Figure 12. The mean naturalness score of the non-aligned utterance was significantly better than both the aligned and the human utterance. It is the only presented option that does not contain any use of the

pragmatic marker "like". The difference between the mean scores of the aligned and the human utterances is not significant. We might conclude that the use of "like" is deemed to be too informal for this scenario, or perhaps too similar in style to the other dialogue partner. A more controlled experiment to study the effects of this feature would be required to make any certain conclusions.

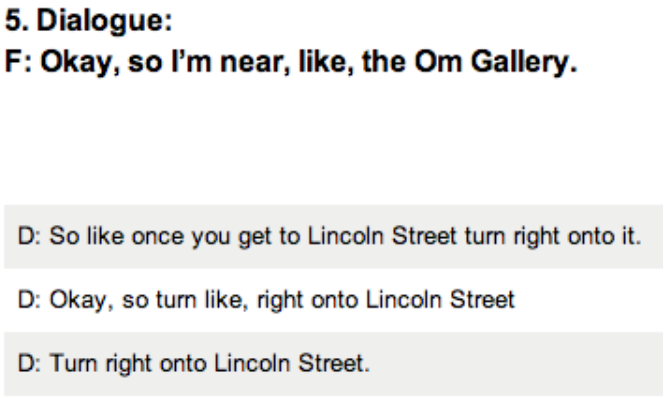


Figure 12: Survey question 5

Question 6 is shown in Figure 13. The multiple comparison test for question 6 revealed that the mean score for the human utterance was significantly different than the means of both the aligned and the non-aligned utterances. The difference between the mean scores of the aligned and non-aligned utterances were not significantly different. The human utterance with a mean score of 2.85 was statistically less natural than the aligned utterance and the non-aligned utterance, with means of 3.66 and 3.89 respectively. A qualitative analysis of the question shows that the human utterance for this question is much more stylized than the other two options. It contains the verbal pause "uh" as well as the phrase "I think". Both of these features can serve as expressions of uncertainty. Because uncertainty is not a highly desirable quality for a director, participants may have perceived this to be an awkward way to present directions and thus assigned it a lower naturalness score. Although this analysis does not reveal anything certain about our generated aligned utterances, we can at least glean that these expressions of uncertainty are potentially undesirable features for language in this domain.

Question 8 is shown in Figure 14. The multiple comparison test for question 8

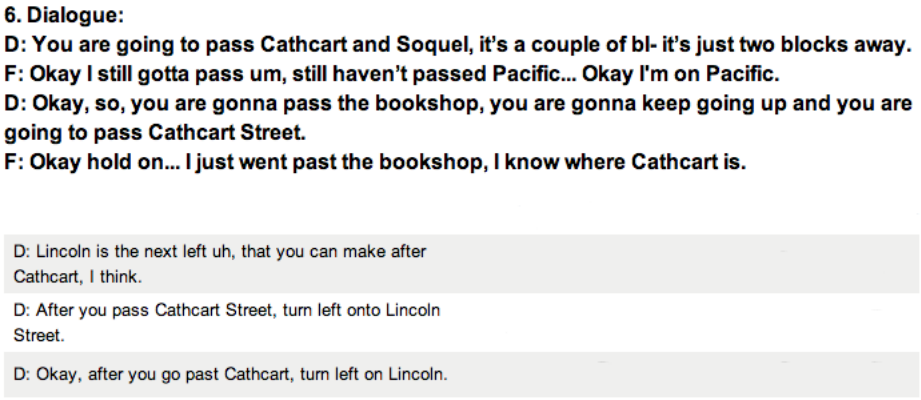


Figure 13: Survey question 6

revealed that the difference between the mean scores of the aligned and the non-aligned utterances was significant. The mean score of the aligned utterance was 3.56 and the mean score of the non-aligned utterance was 4.11. A qualitative evaluation of this question set provides some potentially useful insight. The aligned utterance is perhaps too aligned with the structural and lexical values of the immediately preceding follower utterance, which may give the same mocking effect as postulated for the aligned utterance of question 4. Alignment to the pragmatic marker "okay" is present as well as an alignment of future tense with the lexical feature "will". The previous dialogue turn contains the question, "[W]ill it be on the left or the right side?" When asked a question like this with a binary choice it is common to expect a simple, concise response as either option A or option B. It's possible that the use of "okay" in this context seemed odd, which persuaded participants to regularly select the non-aligned option over the aligned option. This hypothesis would predict that without the preceding "okay", there would be no significant difference in preference between the aligned and the non-aligned. However as observed in question 4, a highly repetitive response can be perceived as mocking and the aligned response does have a significant lexical overlap with the preceding dialogue turn.

In addition to a question-by-question analysis, we looked at the perception of each condition across the entire survey. A multiple comparison test of each condition with all 610 judgements revealed a significant difference between the aligned and non-aligned utterances. That is, the overall mean score of 3.9 for the non-aligned

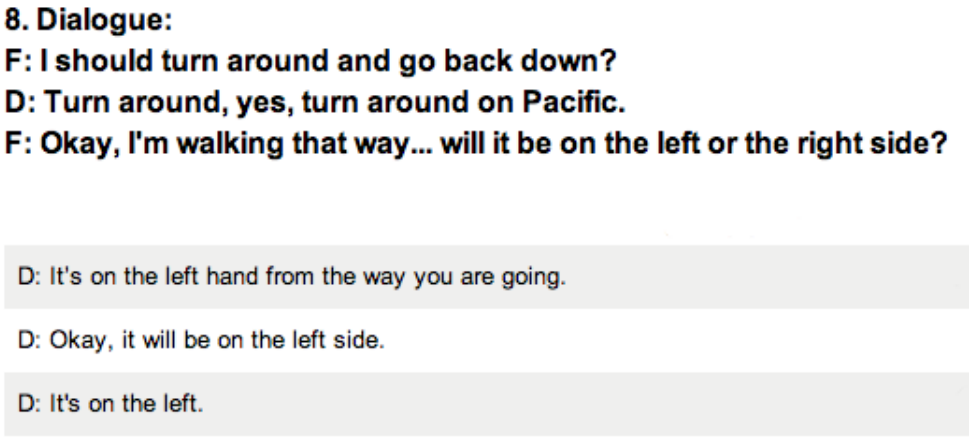


Figure 14: Survey question 8

utterances is significantly better than the overall mean score of 3.5 for the aligned utterances. On a qualitative level, the non-aligned utterances were less stylized and more formal than the aligned utterances. This result provides evidence to support a hypothesis that participants' conversational expectations for this domain are strongly influenced by experiences with existing navigation systems, which is further discussed in the following section.

6 Discussion

The results of the perceptual experiment open a discussion about how various language models can influence language production and comprehension. Participants' ratings could be highly influenced by both personal language preferences and local paradigms. Due to wide variation between individuals and the presence of varied dialects, two speakers of a common language can have significantly different language models and thus significantly different expectations for how a natural conversation would progress. For example, survey question three (shown in Figure 15) included the use of the phrase, "hang a right", which is an alternative expression for the instruction "turn right". Participants who were unfamiliar with this expression likely rated it to be less natural because it was simply inconsistent with their personal language model. Such a bias would contribute negatively to the overall naturalness score. This emphasizes a limitation of our current method and provides a motivation

for extending it to incorporate a more detailed language model of potential users.

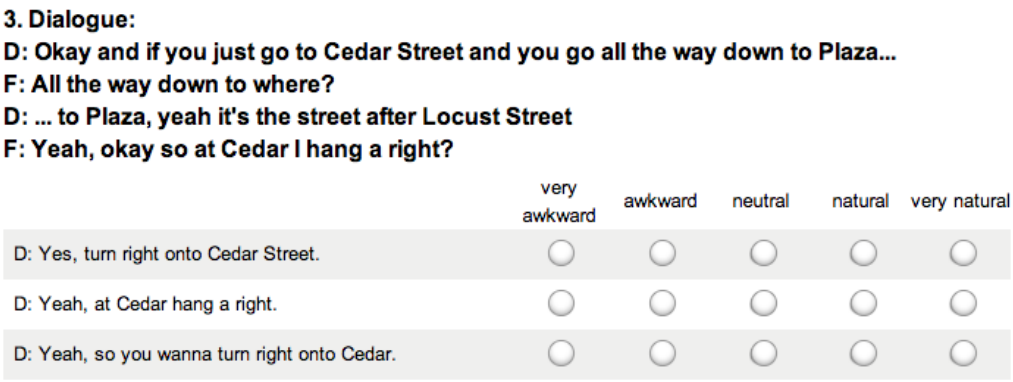


Figure 15: Survey question including the expression "hang a right"

In addition to personal preferences, a speaker's language model may also be strongly influenced by experiences with existing navigation systems that present text to speech turn-by-turn directions, such as TomTom [24] and any of the various navigation applications for cars and mobile phones. The default turn-by-turn directions presented by these systems are generally not very stylized. As such, widespread use of these systems could frame a particular user expectation for how computer generated directions should be presented. This perspective raises an additional question; would strict adherence to an existing paradigm of language reduce the cognitive demands of the dialogue significantly more than alignment does? Further research regarding the effect of alignment in a deployed spoken dialogue system is required to investigate this question more completely.

Another possible explanation for the high ratings of the non-aligned generated utterances may be a result of the experimental design itself. The linguistic patterns of spoken dialogue differ significantly from those of written language. Because the survey was presented as text, participants may have had difficulty imagining the dialogue as a spoken interaction as it was intended to be - pragmatic markers can seem very awkward in text because they are so rarely used in written language. This perspective may explain why the non-aligned utterances were perceived to be more natural than both the aligned utterances and the human utterances, which both included extensive use of pragmatic markers and informal language.

Furthermore, the context in which conversations arise over directions is most often associated with detailed information about an ambiguous area and not necessarily about making left and right turns on well-marked routes. Alignment, therefore, may be more likely to play a role and thus more likely to sound natural when two conversants are attempting to converge toward a mutually understood spatial representation of an area in order to provide directions within it. However, such a system would require extensive and detailed knowledge of an area including landmarks and points of interest, which is a separate research task in and of itself.

7 Conclusion

This paper presents a detailed description and evaluation of PERSONAGE-*primed*, an alignment-capable, paramaterizable natural language generation system that produces lexical, structural and stylistic variation found in natural dialogue. A human evaluation of the output reveals that the aligned utterances do sound natural in the context of relevant dialogue. However, the varied results suggest that a more sophisticated model is necessary in order to reliably elicit the potential benefits of integrating dynamic alignment.

This work provides a springboard for further work on dynamic adaptive language generation. PERSONAGE-*primed* is highly paramaterizable such that individual features may be easily controlled and dispreferred features may be omitted altogether. Because it is capable of producing a large variation of utterances from varied combinations of the existing parameters it can easily be used to support an overgeneration and ranking experiment. Such an experiment would inform future iterations of the system which specific combinations of features are most effective.

In addition, the existing system could be improved in several aspects. The generation dictionary could be autogenerated instead of handcrafted. This would not only save time but would also introduce a method of automatically adapting the model to various domains and languages. The frequency and variation from the existing lexical databases could be integrated into the lexical choice phase in order to better emulate how both existing language models and contextual linguistic

information can influence language production. Psycholinguistic factors such as recency and decay effects could also be integrated into the system to better model the scope of alignment in dialogue.

Finally, the far-reaching goals of this research extend to actual implementations of dynamic alignment in dialogue systems. In order to further understand the real contribution that alignment can make towards improving human-computer interactions, it should be tested in the context of a controlled experiment. If linguistic adaptation is in fact an effective method of improving communication, we predict a measurable increase in satisfaction or a decrease in task duration between interactions with and without alignment. Such future experiments will help evaluate the specific effect of lexical, syntactic, and stylistic alignment on dialogue system interactions.

References

[1] Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25, 2000.

[2] Susan E Brennan and Herbert H Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology-Learning Memory and Cognition*, 22(6):1482–1493, 1996.

[3] Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. Modelling alignment for affective dialogue. In *Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*, 2005.

[4] Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. An alignment-capable microplanner for natural language generation. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 82–89. Association for Computational Linguistics, 2009.

[5] Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In *Proceedings of EMNLP*, volume 4, pages 33–40, 2004.

[6] Herbert H Clark and Jean E Fox Tree. Using¡ i¿ uh¡/i¿ and¡ i¿ um¡/i¿ in spontaneous speaking. *Cognition*, 84(1):73–111, 2002.

[7] Markus De Jong, Mariët Theune, and Dennis Hofs. Politeness and alignment in dialogues with a virtual guide. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pages 207–214. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[8] Christiane Fellbaum. *WordNet*. Springer, 2010.

[9] Jean E Fox Tree and Josef C Schrock. Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, 40(2):280–295, 1999.

[10] Simon Garrod and Anthony Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218, 1987.

[11] Simon Garrod and Martin Pickering. Toward a mechanistic psychology of dialogue: The interactive alignment model. In *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue. BI-DIALOG*, 2001.

[12] Ryuichiro Higashinaka, Rashmi Prasad, and Marilyn A Walker. Learning to generate naturalistic utterances using reviews in spoken dialogue systems. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 265–272. Association for Computational Linguistics, 2006.

[13] Amy Isard, Carsten Brockmann, and Jon Oberlander. Individuality and alignment in generated dialogues. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 25–32. Association for Computational Linguistics, 2006.

[14] Willem JM Levelt and Stephanie Kelter. Surface form and memory in question answering. *Cognitive psychology*, 14(1):78–106, 1982.

[15] Kris Liu, Natalia Blackwell, Jean E. Fox Tree, and Marilyn A. Walker. 21st annual meeting of the society for text and discourse. In *A Hula Hoop almost Hit Me!: Running a Map Task in the Wild to Study Conversational Alignment.*

[16] François Mairesse and Marilyn Walker. Personage: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 496, 2007.

[17] François Mairesse and Marilyn A Walker. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, 20(3):227–278, 2010.

[18] Ani Nenkova, Agustín Gravano, and Julia Hirschberg. High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 169–172. Association for Computational Linguistics, 2008.

[19] Gabriel Parent and Maxine Eskenazi. Lexical entrainment of real users in the lets go spoken dialog system. In *Proceedings Interspeech*, pages 3018–3021, 2010.

[20] Robert Porzel. How computers (should) talk to humans. *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, page 7, 2006.

[21] Antoine Raux, Brian Langner, Dan Bohus, Alan W Black, and Maxine Eskenazi. Lets go public! taking a spoken dialog system to the real world. In *in Proc. of Interspeech 2005*. Citeseer, 2005.

[22] David Reitter and Johanna D Moore. Predicting success in dialogue. In *Annual Meeting-Association for Computational Linguistics*, volume 45, page 808, 2007.

[23] Svetlana Stoyanchev and Amanda Stent. Concept form adaptation in human-computer dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 144–147. Association for Computational Linguistics, 2009.

[24] Portable GPS TomTom. Car navigation systems. *TomTom Navigator*, pages 1–4, 2007.

[25] Arthur Ward and Diane Litman. Dialog convergence and learning. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 158:262, 2007.