Entrainment in Pedestrian Direction Giving: How many kinds of entrainment?

Zhichao Hu, Gabrielle Halberg, Carolynn R. Jimenez and Marilyn A. Walker

Abstract Human conversants in dialog adjust their behavior to their conversational partner in many ways. In terms of language use, they adapt to their partners both lexically and syntactically, by using the same referring expressions or sentence structure. In this paper, we describe a natural language generator PERSONAGE-PRIMED, which can produce utterances entrained to a range of utterance features used in prior utterances by a human user, and represented in the discourse context. PERSONAGE-PRIMED can entrain to the user's referring expressions, tense-modality selection, verb and noun lexical selection, hedge and cue word choice, and syntactic template selection, or any combination of these. To our knowledge, there is no other NLG engine that can dynamically generate all these types of entrainment in any utterance. We report an experiment testing all possible combinations of entrainment in a particular discourse context in order to test whether some types of entrainment are preferred, either because they make the utterance more natural, or because humans perceive the system as more friendly. Our experimental results suggest that human judgements of naturalness are distinct from friendliness: entraining on a user's hedges increase perceptions of friendliness while reducing naturalness, while entraining on user's referring expressions, syntactic template selection and tense/modal choices increase perceptions of both naturalness and friendliness.

1 Introduction

Decades of research on human communication provides substantial evidence that human conversants in dialog adjust their behavior to their conversational partner, but theories differ on whether this results from priming, or beliefs about their partner's knowledge and understanding, or to serve social goals such as communicating liking or to show distance [34, 8, 9, 4, 16, 17, 10, 18, 20, 33, 35, 39, 19]. Conversants lexically entrain or align to particular ways of referring to things [4, 1], and mimic the partner's speech accent, speech rate, pause length, utterance length, and lexical diversity [11, 40]. Humans also entrain to dialog systems, in choice of vocabulary, speech rate and pause length [36, 12, 28]. Research has also shown a positive correlation between measures of entrainment and task success [31, 32, 30, 38, 27]. To date however, the technical challenges of getting a dialog system **to dynamically entrain to the user** has made it difficult to test the potential benefits of user entrainment.

University of California Santa Cruz, Natural Language and Dialog Systems Lab, e-mail: `zhu|crjimen|mawalker@ucsc.edu`

Our work is carried out in the context of the SKIPPER project whose aim is to study “entrainment in the wild” in pedestrian direction giving dialogs. As part of the SKIPPER project, we collected a corpus of human-human dialogs for the pedestrian direction task, the ArtWalk corpus [23]. To our knowledge, this corpus collection is the first experiment to show that entrainment actually occurs in the context of a real task, while people are out in the world, navigating a natural terrain. ArtWalk contains 60 dialogs each with around 450 turns. Every dialog involves a director on campus, and a follower downtown, communicating by phone. The director had access to Google Earth views of the follower’s route and a map, and provided directions to help the follower find public art such as sculptures, mosaics or murals in downtown Santa Cruz. The excerpt of dialog from the ArtWalk corpus in Fig. 1 illustrates entrainment to discourse cues with *okay* in D98 and F98, and in referring expression entrainment in D101 *like a vase*.

Using insights from our analysis of ArtWalk, we developed PERSONAGE-PRIMED, an extension of the PERSONAGE spoken language generator that entrains dynamically to user utterances as represented in the discourse context. The discourse model of PERSONAGE-PRIMED keeps track of user utterance choices in referring expressions, discourse cues, location names, prepositions, and syntactic forms. This allows entrainment to occur at many different stages of the language generation process, and lets PERSONAGE-PRIMED produce tens of different possible utterances in a given context.

Speaker[Utterance #]: Utterance
F97: okay Im on pacific avenue and plaza
D98: okay so you just take a right once your out of pacific lane you go wait no to late to your left.
F98: okay
D99: and I think. its right ther- * alright so im walking down pacific* okay so its right before the object its right before the mission and pacific avenue intersection *okay* its like umm almost brown and kinda like tan colored
F99: is it tan
D100: yeah its like two different colors its like dark brown and orangey kinda like gold color its kinda like um
F100: okay is it kinda like a vase type of a thing
D101: yeah it has yeah like a vase
F101:okay yeah I got it okay one second just take a picture. Alright

Fig. 1 Sample Dialog Excerpt from the ArtWalk Corpus.

In Sec. 2, we first describe the architecture and capabilities of PERSONAGE-PRIMED and the new parameters we developed to make PERSONAGE capable of entraining to the user. Then in Sec. 3, we report on the results of our experiment to test user perceptions of entrainment in human-computer dialog. Our experiment uses an overgenerate and rank paradigm in order to explicitly explore combinations of entrainment parameters and user perceptions of friendliness and naturalness in a particular dialog context. We delay a more detailed comparison of our work to previous work on generating entrainment to Sec. 4 when we can more easily compare it with our own results.

2 Personage Primed Architecture

PERSONAGE-PRIMED is an extension of the parameterizable language generator PERSONAGE [25]. PERSONAGE is capable of producing a wider range of linguistic variation than template-based language generation systems because it dynamically modifies high level representations of the utterances and utilizes external lexical resources including Verbocean [7] and Wordnet [14]. The architecture of PERSONAGE-PRIMED is shown in Fig. 2. We developed PERSONAGE-PRIMED for the pedestrian direction giving domain because our assumption was that “walking around” would be a good context for testing “entrainment in the wild”. Following directions naturally introduces delays between task relevant utterances as the follower navigates an actual landscape. At the same time, pedestrian directions can easily support a range of experimental manipulations. We chose the ArtWalk context, of asking users to find and take pictures of public art because we assumed that there would not be known referring expressions for these artworks, and that we should therefore be able to elicit entrainment to referring expressions, as in earlier work on entrainment. However, we also discovered in the corpus that entrainment seems to occur not just to referring expressions, but also to a whole range of lexical and syntactic choices in dialog. Thus we designed the PERSONAGE-PRIMED generator to have the capability of entraining on any one of these generation choices.

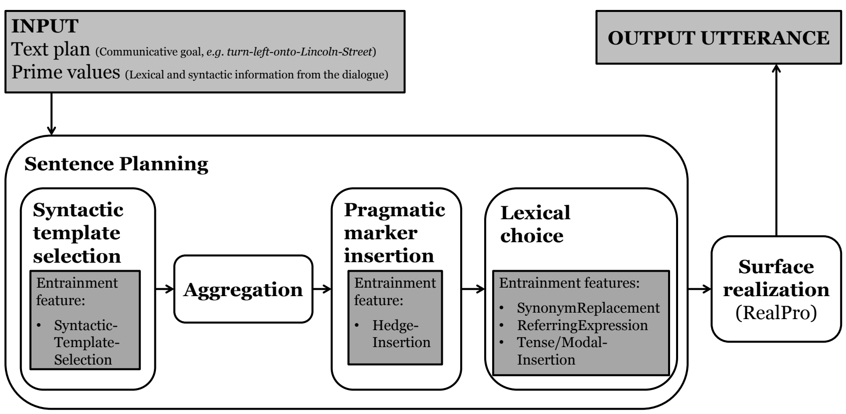

Fig. 2 The architecture of PERSONAGE-PRIMED.

As shown in the architecture diagram in Fig. 2, PERSONAGE supports parameterization of an output utterance via modules for syntactic template selection, pragmatic marker insertion, and lexical choice. In PERSONAGE-PRIMED, the values of the module parameters are controlled by reference to a set of prime values, which represent the content and linguistic information of the dialog context, i.e. the system’s output is generated to entrain with the given dialog. For example, as shown in Fig. 2, lexical choice is further refined into parameters and corresponding prime values for referring expressions, synonyms for nouns and verbs, and the tense/modality to be generated in the system utterance.

Fig. 3 provides an example of how the context is represented by primed values in the discourse model. To our knowledge, PERSONAGE-PRIMED is the first dialog generator to have the capability of entraining on any of the values shown in the discourse model in Fig. 3 and it does so by explicitly manipulating parameters that we have added to PERSONAGE-PRIMED. The prime values contain lexical and syntactic information from the dialog to which the generated utterance will be entrained. An utterance can be produced to entrain to **all** of the entrainment prime values or **none** of them, or **any combination**, depending on the adaptation model in effect when the utterance is dynamically generated by the dialog system. Our goal is to explore which combinations have an effect on user perceptions of system naturalness and friendliness.

Follower: Okay, now I'm at the corner of Cedar Street and Elm, so should I head toward the clock tower from here?
Discourse Context Primed Values:
Prepositions: at, toward, from
Noun: I, corner, here
Tense: present
Modals: should
Verbs: am, head
Place names: cedar street, elm, the clock tower
Hedges: Okay, So, Now
Syntax: (VP PP PP)
Director: confirm + go-to-clocktower

Fig. 3 Sample Discourse Model Representation.

Input. The input to PERSONAGE-PRIMED consists of a text plan and a set of entrainment target values referred to as the prime values as illustrated in Fig. 3. The text plan is a high level semantic representation representing the communicative goal of the desired output utterances. Each text plan contains either a single instruction or a compound instruction. A compound instruction consists of two clauses joined by a temporal relation, such as *after*, *until* or *once*. An example text plan for a compound instruction is shown in Fig. 4. PERSONAGE-PRIMED currently supports 13 unique instructions and statements for the walking directions domain.

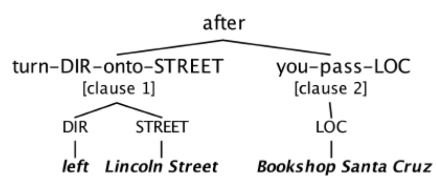

Fig. 4 Example text plan tree.

Syntactic Template Selection. The text plan contains all the information regarding **what** will be communicated, the sentence planning pipeline controls **how** that information is conveyed. Syntactic template selection is the first phase of sentence planning: its goal is to select the most appropriate syntactic form for the instruction(s) in the text plan. Keeping track of user choices in syntactic form is needed in order to produce syntactic entrainment in dialog [3, 2, 31]. If a navigation dialog included the question, *From here where should I go to next?* a

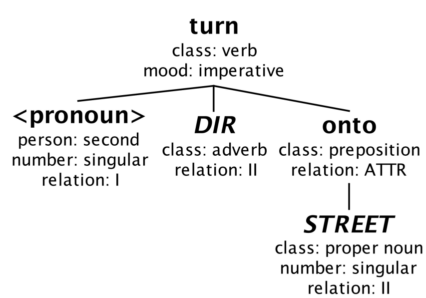

Fig. 5 DSyntS for the instruction *turn-DIR-onto-STREET*. Relation I: the component is the subject of the parent; relation II: the component is the direct object of the parent; relation ATTR: the component is a modifier(adjective/prepositional phrase) of the parent.

response with syntactic entrainment would be phrased in a similar way, such as *From where you are, walk to Pacific Avenue and then make a left.*

PERSONAGE-PRIMED implements the same syntactic dependency tree representation for utterances as used in PERSONAGE [24], referred to as a Deep Syntactic Structure (DSyntS) [22, 26]. The DSyntS specifies the dependency relation between the different components of a sentence. An example DSyntS is shown in Fig. 5. Each instruction and statement has an associated DSyntS List, which is a collection of semantically equivalent DSyntS with different syntactic structure. In order to produce syntactic entrainment, PERSONAGE-PRIMED finds the associated DSyntS List for each instruction in the text plan. It then uses the prime values to select the DSyntS that best matches the lexical and syntactic information. 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 assigned to the instruction.

Aggregation. 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: e.g. **After** *you pass..., turn left onto...* or *Turn left onto...* **after** *you pass....* Currently, there is no entrainment for aggregation operations in PERSONAGE-PRIMED, however in the future, it would be possible to prime particular rhetorical relations and then control the aggregation component as we do other components.

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 serve various pragmatic or social functions. 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 [15, 29]. Discourse markers are especially prevalent in task-oriented dialog.

In PERSONAGE-PRIMED, sample prime values are shown in Fig. 3, e.g. *Okay, Now, So*. The module for pragmatic marker insertion in PERSONAGE-PRIMED will insert up to three of the pragmatic markers found in the prime values.[1] A pragmatic marker is inserted only if one of the insertion points associated with the marker is present in the DSyntS.

Synonym Selection. Synonym selection is a lexical choice operation that 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. See Fig. 2 and the primed context representation in Fig. 3. The system does not currently entrain to nouns because most nouns within the walking directions domain are referring expressions, such as *downtown*, *Pacific Avenue*, etc. Entrainment 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

[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.

like *right* and *left*.

Referring Expression Selection. Referring expression selection is a lexical choice operation that checks 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*}.

This operation also accounts for a referring expression form that is commonly found in navigation dialogs, i.e. referencing street names without the street suffix. If one conversant 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 entrain with this stylistic choice.

Tense transformation and modal insertion. Tense transformation and modal insertion are a final set of lexical choice operations that entrain on primed values for tense and modals. If there exists an explicit use of a particular tense or a modal in the prime values, the current utterance is modified to entrain. The most common tenses used for giving directions in the navigation domain are present, future, and simple future. While followers do use past tense to confirm the completion of an action, it is not common for directors to use it. However, the modals *should*, *can* and *might* are commonly found in navigation dialogs. Followers will express uncertainty with questions such as *Should I stay on Pacific Avenue?*. The corresponding director responses sometimes entrain with this lexical addition with confirming responses such as *Yes, you should stay on Pacific Avenue for three more blocks*.

3 Experimental Method and Results

In a pilot experiment, we asked naive participants from Mechanical Turk to score three utterances in the same context for naturalness: a generated entrained utterance, generated default (non-entrained) utterance and a human utterance which has the same meaning but which is not from the same context. We hypothesized that the entrained generated utterances would be perceived as more natural than the default generated utterances. But the experimental results (default > entrained > human) did not confirm our hypothesis.

In the pilot, every generated entrained utterance entrains to **one or more** of the prime features of its previous (target) utterance, but we did not systematically explore particular parameters (e.g. tense transformation) or combinations (e.g. tense plus cue word). Moreover, there is very little evidence of what people actually do in human-human conversation, and to our knowledge, no previous work has tested whether mimicking all the linguistic features of a conversational partner is natural or whether some kinds of entrainment are dispreferred. Here we aim to systematically explore whether there are clear preferences in types of entrainment by overgenerating possible outputs that entrain on different combinations of prime features. We sample among all the possibilities for entrainment, and our task becomes simply to find out which entrainment combinations are the best. Our earlier work used a similar overgenerate and rank experimental paradigm for collecting data to train a statistical language generator [37, 25].

Ten dialog excerpts are used as context, in which a director (D) is instructing a follower (F) how to navigate to a destination on foot. The dialog excerpts were taken from the Art Walk corpus [23] and were slightly modified to isolate certain priming values. Following the excerpt, participants were presented with options for what the director could say next. Using overgeneration, together with a generated default utterance and a random human utterance, each director response results in 5 to 22 different variations. Having all 22 utterances in one item and asking participants to rank them all does not seem to be a well-defined experimental task. Therefore each item of the experiment survey consists of 5 possible utterances in a particular context, selected so that each possible generated utterance for a particular context appears at least twice across all the survey items. This results in a total of 51 items distributed across 10 surveys. An example item is shown in Fig.6.

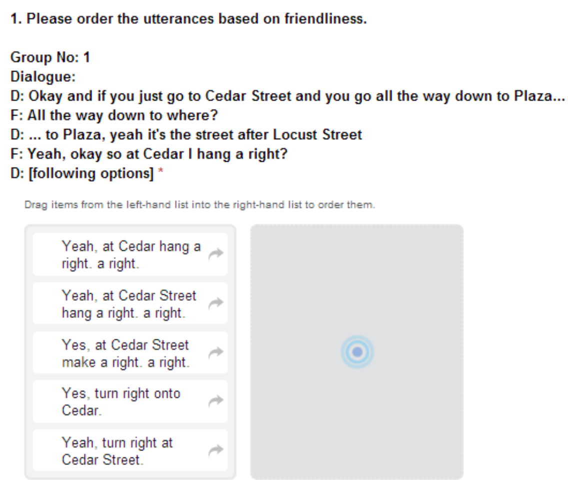

Fig. 6 An example question from Experiment 2.

In one version of the experiment, participants were asked to rank the possible system utterances based on their **naturalness** from high to low. In another version, participants were asked to rank the possible system utterances based on their **friendliness**. This is because default utterances received the highest score for naturalness in the pilot experiment. We hypothesized that one possible explanation of these results was that a director's utterance is considered "natural" when it is concise and clear, and that people may be accustomed to the type of instructions used in current in-vehicle GPS navigation systems. We hypothesized that perceptions of friendliness might be a better probe for entrainment. We hired three judges trained in linguistics to annotate all the combinations we could generate over a period of two weeks, doing two surveys per day at most.

```
Naturalness_score =
  1.6171 * Tense/Modal Insertion +
  1.2486 * ReferringExpression +
  -1.0809 * HedgeInsertion +
  5.5226

Friendliness_score =
  2.9449 * HedgeInsertion +
  2.3239 * RandomHumanUtterance +
  1.4023 * Tense/Modal Insertion +
  0.7223 * SyntacticTemplateSelection +
  3.6261
```

Fig. 7 Regression models.

Feature Value	Meaning
NOMATCH	feature exist AND entrained
MATCHPLUS	feature exist AND not entrained
MATCHMINUS	feature doesn't exist AND not entrained
DEFAULT	generated non-entrained utterance
RAMDOMHUMAN	random human utterance
NULL	for features "Default" and "RandomHumanUtterance", if this feature doesn't exist in the utterance, use NULL

Fig. 8 Possible values for features used in decision tree model.

In order to organize the data for evaluation, we represent the parameters used to generate each of the utterances. Each utterance therefore is represented by 7 features. Five of these are entrainment features: SynonymReplacement, ReferringExpression, HedgeInsertion, SyntacticTemplateSelection and Tense/Modal Insertion as described above in Sec. 2. There were also features representing whether the utterance was a Random Human Utterance or a Default utterance.

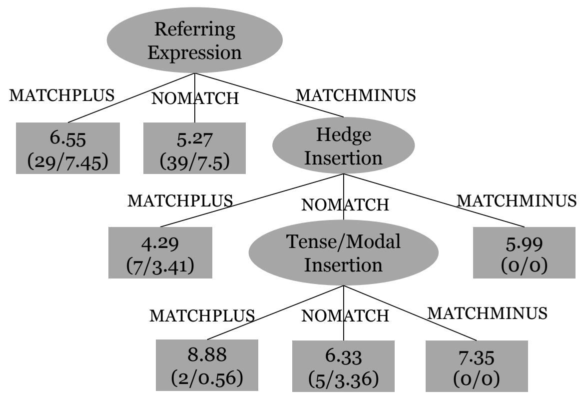

Fig. 9 Decision tree model for Experiment 2 naturalness.

To evaluate the effects of the different parameters, we train two types of models for evaluation: multivariate linear regression models and decision tree models. In the regression model, the dependent variable is the average ranking scores across all three judges across both instances of the utterance. In one ranking question, there are 5 utterances. If an utterance is ranked first in a question, the score of the object is 4. If an utterance is ranked last in a question, the score of the object is 0. We use the sum of the scores from all annotators as the label for an utterance. There are 3 annotators in total, so the scores range from 0 to 12. Since an utterance appears at least twice among all surveys, it will have 2 or more scores. We simply take an average of these scores. The features in regression model are 0/1 features, where a value of 1 indicates that the feature positively affected the generation of the utterance, whereas a value of 0 means this feature was not used in generating the utterance.

We use Linear Regression in Weka 3.6.1 with 10-fold cross validation. Fig. 7 shows the regression models for naturalness and friendliness. In the naturalness model, the correlation coefficient is 0.3055, relative absolute error is 96.332% and root relative squared error is 94.7057%. ReferringExpression and Tense/Modal Insertion both have positive weights, i.e. entraining on these features increases the perception of the naturalness of the utterance. HedgeInsertion has the only negative weight in the model. In contrast, the friendliness model provides a better fit with a correlation coefficient of 0.52, relative absolute error is 81.22% and root relative squared error of 85.76%. Surprisingly, HedgeInsertion has the highest positive weight in the model, suggesting that more hedging leads to perceptions that the system is more friendly. RandomHumanUtterance has the second highest positive weight.

In the decision tree model, the dependent variables are identical to the regression model. However, here we distinguish more values for each features rather than making them binary features. As shown in Figure 8, a feature may have any of 6 possibilities. Recall that there are certain **primed** features represented in the dialog context (10 given dialog excerpts) that the following utterance can entrain to. Since an utterance only entrains to the feature if the feature is primed in context, the combination "feature is entrained AND the context doesn't have feature" cannot occur.

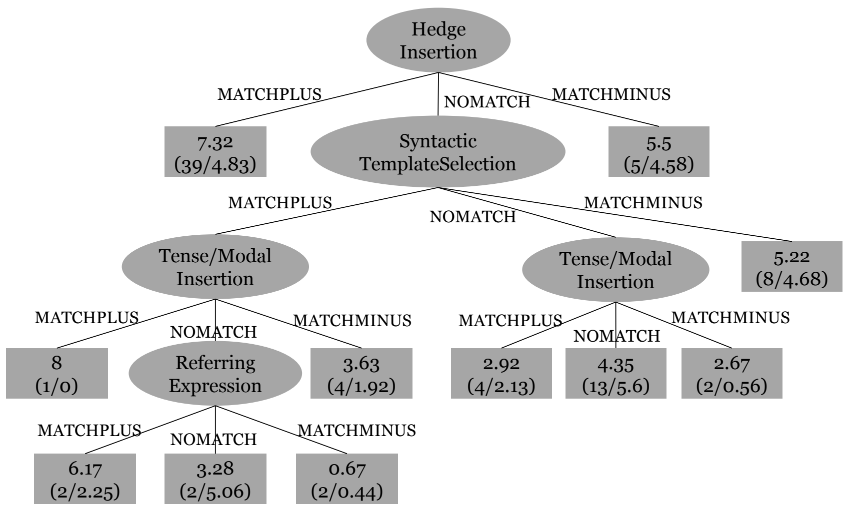

Fig. 10 Decision tree model for Experiment 2 friendliness.

Decision trees are trained using the REPTree package in Weka 3.6.1. We use the whole evaluation data as both training set and test set, and disable pruning to intentionally force the decision tree to overfit the data. Fig. 9 and Fig. 10 show the decision trees for naturalness and friendliness. In the leaf nodes, the first number is the predicted score. The numbers in the parentheses are (number of examples in this leaf / number of misclassified examples on average).

In the naturalness model, the correlation coefficient is 0.35, relative absolute error is 94.81 % and root relative squared error is 93.77%. When previous context provides a prime value for ReferringExpression, and the utterance entrained on the referring expression (MATCHPLUS), we get the highest score with the highest number of examples. If the utterance doesn't entrain on referring expressions (NOMATCH), the scores are relatively lower. If the previous context doesn't provide a prime value for ReferringExpression, then HedgeInsertion primes and utterance features are considered. Similar to the regression model in Fig. 7, hedging is a negative factor in naturalness: utterances that entrain on HedgeInsertion (MATCHPLUS) have a lower naturalness score than utterances that don't (NOMATCH).

In the friendliness model, the correlation coefficient is 0.63, relative absolute error is 74.95% and root relative squared error is 77.50%. These results also indicate that hedging affects perceptions of friendliness as in the regression model shown in Fig. 7. When the dialog context provides a prime value for HedgeInsertion, and the utterance entrains for hedging (MATCHPLUS), the resulting friendliness score is the highest with the highest number of examples. If the utterance did not entrain on hedging (NOMATCH) even though a prime value was available, then SyntacticTemplateSelection is considered by the model. Generally, utterances that entrain on SyntacticTemplate Selection have higher scores.

4 Discussion and Conclusion

This paper presents an experiment based on PERSONAGE-PRIMED, an extended version of PERSONAGE that can dynamically entrain to the dialog context. We show that some types of entrainment have a positive effect on the friendliness of system utterances, while other types positively effect perceptions of naturalness.

Previous work testing the benefits of entrainment have been measured in different contexts, such as whether entrainment in human-human dialog predicts success. Much of the previous work on human-computer dialog has examined whether the human entrained to the computer rather than vice versa. Our work contributes to the limited amount of previous work on adaptive generation using different computational methods for generation. Jong et al. [13] present an approach that focuses on affective language use for aligning specifically to user's politeness and formality. Brockman et al.'s model [5] simulates alignment using word sequences alone. An extension of this work in Isard et al. [21] simulates both individuality and alignment in dialog 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.

Perhaps most similar to our goals is the alignment-capable microplanner SPUD *prime* presented by Buschmeier et al [6]. SPUD *prime* is a computational model for language generation in dialog 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 dialog. 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 dialog from a controlled experiment.

In a field study conducted with the Let's Go system however, user utterance behavior was batched to produce new system behaviors in a non-dynamic version of the system, but which however produced behaviors entrained to user behavior in the corpus collected earlier. This study showed that system entrainment to the user could be helpful, but the switch in system behavior may have confused some users. In contrast, we test a system that is capable of dynamic entrainment, but we test it in the lab with user perceptions. While this is the first study to our knowledge to be based on a generator that can produce utterances dynamically entrained to any primed feature or combination of primed features in the context, in future work, we hope to be able to test dynamically produced entrainment in the field.

Acknowledgements This research was supported by NSF CISE RI EAGER #IIS-1044693 and NSF CISE CreativeIT #IIS-1002921, and auxiliary REU supplements. The ArtWalk corpus was collected in collaboration with Prof. Jean E. Fox Tree, UCSC.

References

1. Branigan, H., Pickering, M., Pearson, J., McLean, J.: Linguistic alignment between people and computers. Journal of Pragmatics **42**(9), 2355–2368 (2010)
2. Branigan, H., Pickering, M., Pearson, J., McLean, J., Nass, C.: Syntactic alignment between computers and people: the role of belief about mental states. In: Proceedings of the 25th Annual Conference of the Cognitive Science Society, pp. 186–191 (2003)
3. Branigan, H.P., Pickering, M.J., Cleland, A.A.: Syntactic co-ordination in dialogue. Cognition **75**(2), B13–B25 (2000)
4. Brennan, S.E., Clark, H.H.: Lexical choice and conceptual pacts in conversation. Journal of Experimental Psychology: Learning, Memory And Cognition (1996)
5. Brockmann, C., Isard, A., Oberlander, J., White, M.: 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)
6. Buschmeier, H., Bergmann, K., Kopp, S.: An alignment-capable microplanner for natural language generation. In: Proceedings of the 12th European Workshop on Natural Language Generation, pp. 82–89. Association for Computational Linguistics (2009)
7. Chklovski, T., Pantel, P.: Verbocean: Mining the web for fine-grained semantic verb relations. In: Proceedings of EMNLP, vol. 4, pp. 33–40 (2004)
8. Clark, H.H.: Using Language. Cambridge University Press (1996)
9. Clark, H.H., Brennan, S.E.: Grounding in communication. In: L.B. Resnick, J. Levine, S.D. Bahrend (eds.) Perspectives on socially shared cognition. APA (1991)
10. Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition **22**, 1–39 (1986)
11. Coupland, N., Coupland, J., Giles, H., Henwood, K.: Accommodating the elderly: Invoking and extending a theory. Language in Society **17**(1), 1–41 (1988)
12. Darves, C., Oviatt, S.: Adaptation of users' spoken dialogue patterns in a conversational interface. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP'2002) (2002)
13. De Jong, M., Theune, M., Hofs, D.: 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, pp. 207–214. International Foundation for Autonomous Agents and Multiagent Systems (2008)
14. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)
15. Fox Tree, J., Schrock, J.: Discourse Markers in Spontaneous Speech: Oh What a Difference an Oh Makes. Journal of Memory and Language **40**(2), 280–295 (1999)
16. Fox Tree, J., Schrock, J.: Basic meanings of you know and I mean. Journal of Pragmatics **34**(6), 727–747 (2002)
17. Fox Tree, J.E.: Listening in on monologues and dialogues. Discourse Processes **27**, 35–53 (1999)
18. Fox Tree, J.E., Mayer, S.A.: Overhearing single and multiple perspectives. Discourse Processes **45**(160-179) (2008)
19. Giles, H., Coupland, N.: Ethnicity and intergroup communication pp. 21–42 (1991)
20. Isaacs, E.A., Clark, H.: References in conversations between experts and novices. Journal of Experimental Psychology **116**(1), 26–37 (1987)
21. Isard, A., Brockmann, C., Oberlander, J.: Individuality and alignment in generated dialogues. In: Proceedings of the Fourth International Natural Language Generation Conference, pp. 25–32. Association for Computational Linguistics (2006)
22. Lavoie, B., Rambow, O.: A framework for customizable generation of multi-modal presentations. In: COLING-ACL98. ACL, Montréal, Canada (1998)
23. Liu, K., Blackwell, N., Tree, J.E.F., Walker, M.A.: 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
24. Mairesse, F., Walker, M.: Towards personality-based user adaptation: psychologically informed stylistic language generation. User Modeling and User-Adapted Interaction pp. 1–52 (2010)
25. Mairesse, F., Walker, M.A.: Controlling user perceptions of linguistic style: Trainable generation of personality traits. Computational Linguistics (2011)
26. Melčuk, I.A.: Dependency Syntax: Theory and Practice. SUNY, Albany, New York (1988)

27. Nenkova, A., Gravano, A., Hirschberg, J.: 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, pp. 169–172. Association for Computational Linguistics (2008)
28. Parent, G., Eskenazi, M.: Lexical entrainment of real users in the lets go spoken dialog system. In: Proceedings Interspeech, pp. 3018–3021 (2010)
29. Passonneau, R.J., Litman, D.: Empirical analysis of three dimensions of spoken discourse: Segmentation, coherence and linguistic devices. In: D. Scott, E. Hovy (eds.) Computational and Conversational Discourse: Burning Issues - An Interdisciplinary Account, pp. 161–194. Springer-Verlag, Heidelberg, Germany (1996)
30. Porzel, R.: How computers (should) talk to humans. How People Talk to Computers, Robots, and Other Artificial Communication Partners p. 7 (2006)
31. Reitter, D., Keller, F., Moore, J.: Computational modelling of structural priming in dialogue. Proc. Human Language Technology conference-North American chapter of the Association for Computational Linguistics annual mtg (2006)
32. Reitter, D., Moore, J.: Predicting Success in Dialogue. Annual Meeting-Association for Computational Linguistics **45**(1), 808 (2007)
33. Schober, M.F.: Spatial perspective-taking in conversation. Cognition **47**, 1–24 (1993)
34. Schober, M.F.: Different kinds of conversational perspective-taking. In: S.R. Fussell, R.J. Kreuz (eds.) Social and cognitive psychological approaches to interpersonal communication, pp. 145–174. Lawrence Erlbaum (1998)
35. Schober, M.F., Clark, H.H.: Understanding by addressees and overhearers. Cognitive Psychology **21**(2), 211–232 (1989)
36. Stoyanchev, S., Stent, A.: 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, pp. 144–147. Association for Computational Linguistics (2009)
37. Walker, M.A., Stent, A., Mairesse, F., Prasad, R.: Individual and domain adaptation in sentence planning for dialogue. Journal of Artificial Intelligence Research (JAIR) **30**, 413–456 (2007)
38. Ward, A., Litman, D.: Dialog convergence and learning. Frontiers in Artificial Intelligence and Applications **158**, 262 (2007)
39. Wilkes-Gibbs, D., Clark, H.H.: Coordinating beliefs in conversation. Journal of Memory and Language **31**, 183–194 (1992)
40. Willemyns, M., Gallois, C., Callan, V.J., Pittam, J.: Accent accommodation in the job interview impact of interviewer accent and gender. Journal of Language and Social Psychology **16**(1), 3–22 (1997)