Individuality and Alignment in Generated Dialogues

Amy Isard and **Carsten Brockmann** and **Jon Oberlander**
School of Informatics, University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW, UK
{Amy.Isard, Carsten.Brockmann, J.Oberlander}@ed.ac.uk

Abstract

It would be useful to enable dialogue agents to project, through linguistic means, their individuality or personality. Equally, each member of a pair of agents ought to adjust its language (to a greater or lesser extent) to match that of its interlocutor. We describe CRAG, which generates dialogues between pairs of agents, who are linguistically distinguishable, but able to align. CRAG-2 makes use of OPENCCG and an over-generation and ranking approach, guided by a set of language models covering both personality and alignment. We illustrate with examples of output, and briefly note results from user studies with the earlier CRAG-1, indicating how CRAG-2 will be further evaluated. Related work is discussed, along with current limitations and future directions.

1 Introduction

A computer agent should be *individual*. Nass and collaborators find that users' responses to computer-agents are influenced by whether the agent's linguistic personality matches—or mismatches—the personality of the user (Moon and Nass, 1996; Nass and Lee, 2000). Similarly, characters in virtual environments should be distinctive (Ball and Breese, 2000; Rist et al., 2003). But an aspect of personality is how well you adjust to other people (and their language use): *alignment*. Pickering and Garrod's Interactive Alignment Model suggests that people tend to automatically converge on lexical and syntactic choices, via a low-level mechanism of interpersonal priming (Pickering and Garrod, 2004), and Brennan has shown that people will align their language towards that of computer agents (Brennan, 1996). But it is an open issue as to whether some people are better 'aligners' than others. Conversely, alignment is only visible and interesting (among computer agents) if they start out being individual.

We therefore set out to simulate *both* individuality and alignment. The paper briefly surveys the evidence for linguistic personality, for interpersonal alignment, and for interaction between them. It then sketches the current version of CRAG. CRAG-2 makes use of OPENCCG and an over-generation and ranking approach, guided by a set of language models for personality and alignment. We illustrate the differing linguistic behaviours that it generates, and briefly note promising results from user studies with the earlier CRAG-1 system, indicating how CRAG-2 will be further evaluated. Related work is discussed, along with possible directions for future work.

2 Background

2.1 Personality and Language

Current work on personality traits is dominated by Costa and McCrae's five-factor model (Costa and McCrae, 1992). The five factors, or dimensions, are: Extraversion; Neuroticism; Openness; Agreeableness; and Conscientiousness (Matthews et al., 2003). It has been shown that scores on these dimensions correlate with some aspects of language use (Scherer, 1979; Dewaele and Furnham, 1999). In studies of text, the focus has been on lexical choice, and Pennebaker and colleagues have analysed relative frequencies of use of word-stems in a dictionary structured into semantic and syntactic categories (Pennebaker et al., 2001). Amongst other results, they have shown that High Extraverts

Proceedings of the Fourth International Natural Language Generation Conference, pages 25–32,
Sydney, July 2006.

use: more social process talk, positive emotion words and inclusives; and fewer negations, tentative words, exclusives, causation words, negative emotion words, and articles (Pennebaker and King, 1999; Pennebaker et al., 2002).

Computational linguistic exploitation of such empirically-derived features has been limited. On the one hand, in generation, there has been work on personality-based generation. For instance, in developing embodied conversational agents, researchers have designed agents or teams of agents with distinguishable linguistic personalities (Ball and Breese, 2000; Rist et al., 2003; Piwek and van Deemter, 2003; Gebhard, 2005). However, the linguistic behaviour is usually informed by rules based on personality stereotypes, rather than on language statistics themselves. On the other hand, in interpretation, more empirical work has recently been carried out, to enable text classification. Argamon et al. (2005) attempted to classify authors as High or Low Extravert and High or Low Neurotic, using Pennebaker and King's (1999) data. They report classification accuracies of around 58% (with a 50% baseline). Oberlander and Nowson (2006) undertake a comparable task, using weblog data. They report classification accuracies of roughly 85% (Neuroticism) and 94% (Extraversion), and comparable figures for Agreeableness and Conscientiousness. Such studies can provide ordered lists of linguistic features which are useful for distinguishing language producers, and we will return to this, below.

2.2 Alignment and Language

People converge with their interlocutors in linguistic choices at a number of levels (Pickering and Garrod, 2004). The phenomena can be seen in both social and cognitive terms. On the social side, co-operative processes such as audience design are usually considered to be conscious, at least in part (Bell, 1984). But on the cognitive side, co-ordinative processes such as alignment are usually considered to be largely automatic (Garrod and Doherty, 1994). Alignment can be probed by psycholinguistic tests for interpersonal priming, establishing the extent to which participants are more likely to use a lexical item or syntactic construction after hearing their conversational partner use it. Syntactic priming experiments involve constructions such as passives, and ditransitives (Pickering and Branigan, 1998).

It is possible that some people are stronger aligners than others. Gill et al. (2004) probed syntactic priming for passives, and investigated whether levels of Extraversion or Neuroticism would affect the strength of priming effects. It was found that Extraversion has no effect, but that Neuroticism has a non-linear effect: both High and Low levels of Neuroticism led to weaker priming; Mid levels led to significantly stronger priming. Given this, if a generation system is going to simulate alignment, it is probably worth designing it so that it can simulate agents with differing propensities to align.

3 The CRAG System Overview

The system described in the following sections (CRAG-2) is the successor to CRAG-1 which is detailed in Isard et al. (2005). The system generates a dialogue between two computer agents on the subject of opinions about a film. CRAG-2 uses the OPENCCG parsing and generation framework (White, 2004; White, 2006). The realiser component takes a logical form as input and outputs a list of candidate sentences ranked using one or more language models. In CRAG-2, we use the OPENCCG generator to massively over-generate paraphrases, and the combination of n-gram models described in Section 4 to choose the best utterance according to a character's personality and agenda, and the dialogue history.

4 N-Grams: Personality and Alignment Modelling

4.1 N-Gram Language Models

The basic assumption underlying CRAG-2 is that personality, as well as alignment behaviour, can be modelled by the combination of a variety of n-gram language models.

Language models are trained on a corpus and subsequently used to compute probability scores of word sequences. An n-gram language model approximates the probability of a word given its history of the preceding $n-1$ words. According to the chain rule, probabilities are then combined by multiplication. Equation (1) shows a trigram model that takes into account two words of context to predict the probability of a word sequence w_1^n:

$$P(w_1^n) \approx \prod_{i=1}^{n} P(w_i | w_{i-2}^{i-1}) \quad (1)$$

4.2 Avoiding the Length Effect

Because word probabilities are always less than 1 and therefore each multiplication decreases the total, if we use this standard model, longer sentences will always receive lower scores (this is known as the length effect). We therefore calculate the probability of a sentence as the geometric mean of the probability of each word in the sentence as shown in (2):

$$P(w_1^n) \approx \prod_{i=1}^{n} P(w_i|w_{i-2}^{i-1})^{1/n} \quad (2)$$

4.3 Linear Combination of Language Models

OPENCCG supports the linear combination of language models, where each model is assigned a weight. For uniform interpolation of two language models P_a and P_b, each receives equal weight:

$$P(w_i|w_{i-2}^{i-1}) = \frac{P_a(w_i|w_{i-2}^{i-1}) + P_b(w_i|w_{i-2}^{i-1})}{2} \quad (3)$$

In the more general case, the language models are assigned weights λ_i, the sum of which has to be 1:

$$P(w_i|w_{i-2}^{i-1}) = \lambda_1 P_{\mathrm{a}}(w_i|w_{i-2}^{i-1}) + \lambda_2 P_{\mathrm{b}}(w_i|w_{i-2}^{i-1}) \quad (4)$$

For example, setting $\lambda_1 = 0.9$ and $\lambda_2 = 0.1$ assigns a high weight to the first language model.

4.4 OPENCCG N-Gram Ranking

In the OPENCCG framework, language models can be used to influence the chart-based realisation process. The agenda of edges is re-sorted according to the score an edge receives with respect to a language model. For CRAG-2, many paraphrases are generated from a given logical form, and they are then ranked in order of probability according to the combination of n-gram models appropriate for the character and stage of the dialogue.

5 CRAG-2 Personality and Alignment Models

We use the SRILM toolkit (Stolcke, 2002) to compute our language models. All models (except for the cache language model described in Section 5.4) are trigram models with backoff to bigrams and unigrams.

We have experimented with two strategies for creating personality models. Since we want to study the effects of alignment as well as personality, it is essential that the two characters in a dialogue be distinct from one another, so that the effects of alignment can be seen. The first strategy involves using typical language for each personality trait, and the second uses the language of one individual. In both cases, the language models described in the following sections are combined as described in Section 5.5.

5.1 Building a Personality

Nowson (2006) performed a study on language use in weblogs. The weblog authors were asked to complete personality questionnaires based on the five-factor model (see Section 2.1). All weblog authors scored High or Medium on the Openness dimension, so we have no data for typical Low Open language.

We divided the data into High, Medium and Low for each personality dimension, and trained language models so that we would be able to assess the probability of a word sequence given a personality type. This means that each individual weblog is used 5 times, once for each dimension.

For each personality dimension, the system simplifies a character's personality setting x by assigning a value of High ($x > 70$), Medium ($30 < x \leq 70$) or Low ($x \leq 30$). The five models corresponding to the character's assigned personality are uniformly interpolated to give the final personality model. If the character has been given a low Openness score, since we do not have a model for this personality type, we simply interpolate the other four models.

5.2 Borrowing a Personality

Our second strategy was to train n-gram models on language of the individuals from the CRAG-1 corpus (Isard et al., 2005) and to use one of these models for each character in the dialogue.

5.3 Base Language Model

In the case of building a personality, a base language model is obtained by combining a language model computed from the corpus collected for the CRAG-1 system and a general language model based on data from the Switchboard corpus (Stolcke et al., 2000). The combined base model alone would rank the utterances without any bias for personality or alignment. When we are borrowing a personality, the base model is calculated from the Switchboard corpus alone.

5.4 Cache Language Model

We simulate alignment by computing a cache language model based on the utterance that was generated immediately before. This dialogue history cache model is the uniform interpolation of word- and class-based n-gram models, where classes act as a backoff mechanism when there is no exact word match. Classes group together lexical items with similar semantic properties, e.g.:

- *good*, *bad*: `quality-adjective`
- *loved*, *hated*: `opinion-verb`

Details of this approach can be found in Brockmann et al. (2005).

5.5 Combining the Language Models

The system uses weights to combine all the models described above. First the base and personality models are interpolated to produce a base-personality model, and finally the cache model is introduced to add alignment effects.

6 Dialogue and Utterance Specifications

6.1 Character Specification

Two computer characters are parameterised for their personality by specifying values (on a scale from 0 to 100) for the five dimensions: Extraversion (E), Neuroticism (N), Openness (O), Agreeableness (A), and Conscientiousness (C). Their alignment behaviour is set to a value between 0 (low propensity to align) and 1 (high propensity to align). Also, each character receives an agenda of topics they wish to discuss, along with polarities (positive/negative) that indicate their opinion on the respective topic.

6.2 Utterance Design

The character with the higher Extraversion score begins the dialogue, and their first topic is selected. Once an utterance has been generated, the other character is selected, and the system applies the algorithm shown in (5) to decide which topic should come next. This process continues until there are no topics left on the agenda of the current speaker.

(5) **if** (A $<$ 46) or (C $<$ 46) or
(no. of utts about this topic $=$ 2)
then take next topic from own agenda
else continue on same topic

The system creates a simple XML representation of the character's utterance, using the specified topic and polarity. An example using the topic music and polarity negative is shown in Figure 1. At this point the system also decides which discourse connectives may be appropriate, based on the previous topic and polarity.

```
<utterance>
  <utt topic="music" polarity="dislike"
       opp-polarity="like" so="no" right="no"
       also="no" well="yes" and="no" but="no">
    <pred adj="bad"/>
    <opp-pred adj="good"/>
  </utt>
</utterance>
```

Figure 1: Simple Utterance Specification

6.3 OpenCCG Logical Forms

Following the method described in Foster and White (2004), the basic utterance specification is transformed, using stylesheets written in the XSL transformation language, into an OpenCCG logical form. We make use of the facility for defining optional and alternative inputs and underspecified semantics to massively over-generate candidate utterances. A fragment of the logical form which results from the transformation of Figure 1 is shown in Figure 2. We also include some fragments of canned text from the CRAG corpus in our OpenCCG lexicon.

We also add optional interjections (*i mean*, *you know*, *sort of*) and conversational markers (*right*, *but*, *and*, *well*) where appropriate given the discourse history.

When the full logical form is processed by the OpenCCG system, the output consists of sentences of the types shown below:

(I think) the music was bad.
(I think) the music was not (wasn't) good.
I did not (didn't) like the music.
I hated the music.
One thing I did not (didn't) like was the music.
One thing I hated was the music.

The fragmentary logical form in Figure 2 would create all possible paraphrases from:

(well) (you know) I (kind of) [liked/loved] the [music/score]

By using synonyms (e.g., plot=story, comedy=humour) and combining the sentence types

```
<node id="l1:opinion" pred="like" tense="past">
  <rel name="Speaker">
    <node id="p1:person" pred="pro1" num="sg"/>
  </rel>
  <rel name="Content">
    <node id="f1:cragtopic" pred="music"
          det="the" num="sg"/>
  </rel>
  <opt>
    <rel name="Modifier">
      <node id="w1:adv" pred="well"/>
    </rel>
  <opt>
  <opt>
    <rel name="HasProp">
      <node id="a2:proposition" pred="kind-of"/>
    </rel>
  </opt>
  <opt>
    <rel name="Modifier">
      <node id="a1:adv" pred="you-know"/>
    </rel>
  </opt>
</node>
```

Figure 2: Fragment of Logical Form

Stan: E:53 N:48 A:57 C:46 O:65
agenda: film(neg), dialogue(neg), music(pos)
other opinions: plot(neg), comedy(neg)

Eddie: E:51 N:43 A:57 C:41 O:65
agenda: plot(neg), comedy(neg), dialogue(neg)
other opinions: music(pos), film(neg)

Figure 3: Stan and Eddie

and optional expressions, we create up to 3000 possibilities per utterance, and the best candidate is chosen by the specific combination of n-gram models appropriate for the given personality and dialogue history, as described in Section 4.

Our OPENCCG lexicon is based on the core English lexicon included with the system and we have added vocabulary appropriate to the movie domain, and extended the range of grammatical constructions where necessary.

7 Output and Evaluation

7.1 Output

In this section, we provide some example outputs from the CRAG-2 system, using characters based on participants from our corpus (see Section 5.2). Stan is higher on the Extraversion, Neuroticism, and Conscientiousness scales than Eddie. The characters' personalities and agendas are summarised in Figure 3.

We show three example dialogues between Stan and Eddie. In the first (Figure 4) neither character aligns with the other at all, while in the second (Figure 5) Stan has a slight tendency towards alignment and in the third (Figure 6) a more pronounced tendency. In system terms, this means that in the first dialogue the cache model was given weight 0, while in the second and third the cache model was given weights 0.05 and 0.1 respectively for Stan's utterances. It can be seen that in the first dialogue, Eddie's use of *humour* in utterance 4 is followed by Stan's *comedy*, and *dialogue* in utterance 6 is followed by *script*. In the second dialogue, Stan aligns his first lexical choice with Eddie's but not his second, while in the third dialogue he aligns both. The syntactic structures of Stan's utterances in the second and third dialogue also mirror Eddie's progressively more closely.

1. Stan: I didn't think much of the film you know.
2. Eddie: I mean I didn't like kind of like the story.
3. Stan: Yeah there were so many little pathways that they didn't take you know.
4. Eddie: I mean I didn't think the humour was really funny.
5. Stan: Yeah you know the comedy was a bit dull.
6. Eddie: I mean I didn't like kind of like the dialogue.
7. Stan: I mean the script was a bit dull.

Figure 4: Zero Alignment

1. Stan: I didn't think much of the film you know.
2. Eddie: I mean I didn't like kind of like the story.
3. Stan: Yeah there were so many little pathways that they didn't take you know.
4. Eddie: I mean I didn't think the humour was really funny.
5. Stan: I mean the humour was a bit dull.
6. Eddie: I mean I didn't like kind of like the dialogue.
7. Stan: I mean the script was a bit dull.

Figure 5: Little Alignment from Stan

1. Stan: I didn't think much of the film you know.
2. Eddie: I mean I didn't like kind of like the story.
3. Stan: I mean the story was a bit dull.
4. Eddie: I mean I didn't think the humour was really funny.
5. Stan: I mean the humour was a bit dull.
6. Eddie: I mean I didn't like kind of like the dialogue.
7. Stan: I mean the dialogue was a bit dull.

Figure 6: More Alignment from Stan

To further illustrate the differences between the dialogues with and without alignment, we provide some utterance rankings. We show candidates for the fifth utterance in each dialogue. Table 1 shows sentences from the example generated without alignment, corresponding to utterance 5 (Stan)

1	.03317	Yeah you know the comedy was a bit dull.
3	.03210	Yeah you know the humour was a bit dull.
6	.03083	Yeah to be honest I didn't think that the comedy was very good either.
15	.02938	I didn't think much of the comedy either.
24	.02861	I thought that the comedy was a bit dull too you know.

Table 1: Ranked Sentences with Zero Alignment

1	.05384	I mean the humour was a bit dull.
8	.05239	The humour wasn't really funny you know.
15	.04748	I mean I didn't think that the humour was very good either.
19	.04518	I didn't think much of the humour either you know.
21	.04478	I thought the humour was a bit dull too you know.

Table 2: Ranked Sentences with Little Alignment from Stan

from Figure 4. We show the first five occurrences of different sentence structures (see Section 6.3), with their rank and their geometric mean adjusted scores.

Table 2 shows the the top five sentences from the fifth utterance from Figure 5 (little alignment), and Table 3 those from Figure 6 (more alignment). It can be seen that when more alignment is present, the syntactic structure used by the previous speaker rises higher in the rankings.

7.2 Evaluation

We have not evaluated CRAG-2. However, we have evaluated CRAG-1. The method was to generate a set of dialogues, systematically contrasting characters with extreme settings for the personality dimensions (High/Low Extraversion, Neuroticism, and Psychoticism[1]).

[1] CRAG-1 used the simpler PEN three factor personality model.

1	.07081	I mean the humour was a bit dull.
2	.06432	The humour wasn't really funny you know.
15	.05516	I mean I didn't think that the humour was really funny either.
27	.05000	I thought the humour was a bit dull too you know.
36	.04884	I mean I didn't think much of the humour either.

Table 3: Ranked Sentences with More Alignment from Stan

Human subjects were asked to fill in a questionnaire to determine their personality. They were then given a selection of dialogues to read. After each dialogue, they were asked to rate their perception of the interaction and of the characters involved by assigning scores to a number of adjectives related to the personality dimensions.

It was found that subjects could recognise differences in the Extraversion level of the language. Also, the personality setting of a character influenced the perception of its and its dialogue partner's personality (Kahn, 2006).

We plan a similar evaluation for CRAG-2 to be able to compare human raters' impressions of dialogues generated by the two systems. We also plan to evaluate CRAG-2 internally by varying the weight given to the underlying language models, and observing the effects this has on the resulting ranking of the generated utterances.

8 Related Work

Related work in NLG involves either personality or alignment. So far as we can tell, there is little work on the latter. Varges (2005) suggests that "a word similarity-based ranker could align the generation output (i.e. the highest-ranked candidate) with previous utterances in the discourse context", but there is no report yet on an implementation of this proposal. A rather different approach is suggested by Bateman and Paris (2005), who discuss initial work on alignment, mediated by a process of register-recognition. Regarding generation with personality, the most influential work is probably Hovy's PAULINE system, which varies both content selection and realisation according to an individual speaker's goals and attitudes (Hovy, 1990). In her extremely useful survey of work on affective (particularly, emotional) natural language generation, Belz (2003) notes that the complexity of PAULINE's rule system means that numerous rule interactions can lead to unpredictable side effects. In response, Paiva and Evans (2004) take a more empirical line on style generation, which is closer to that pursued here. Other relevant work includes Loyall and Bates (1997), who explicitly propose that personality and emotion could be used in generation, but Belz observes that technical descriptions of Hap and the Oz project suggest that the proposals were not implemented. Walker et al.'s (1997) system produces linguistic behaviour which is much more varied than our current sys-

tem is capable of; but there, variation is driven by a model of social relations (based on Brown and Levinson), rather than on personality. The NECA project subsequently developed methods for generating scripts for pairs of dialogue agents (Piwek and van Deemter, 2003), supported by the MIAU platform (Rist et al., 2003). The VIRTUALHUMAN project is a logical successor to this work, and its ALMA platform provides an integrated approach to affective generation, covering emotion, mood and personality (Gebhard, 2005).

9 Conclusion and Next Steps

Our current system takes a much coarser-grained approach to semantics and discourse goals than the recent projects described above, in order to take advantage of empirically-derived relations between language and personality. It should be feasible in principle to move to a more sophisticated semantics, but still retain the massive overgeneration and ranking method. However, to support more perceptible variation, we need to exploit much larger personality-corpus resources than have been available up to now, and our current priority is to obtain a corpus at least an order of magnitude larger than what is currently available. This interest in individual differences and what corpora can (and cannot) tell us about them is one we share with Reiter and colleagues (Reiter and Sripada, 2004).

We also plan to integrate techniques from CRAG-1 and CRAG-2, by passing the ranked output of CRAG-2 through further processing and ranking stages. Furthermore, we intend to investigate longer-ranging alignment processes, taking into account more than one previous utterance, with reduced weight by distance, to emulate memory effects.

With these enhancements, we will take further steps towards our goal of simulating both individuality and alignment in believable computer agents.

10 Acknowledgements

This research has been funded by Scottish Enterprise through the Edinburgh-Stanford Link project "Critical Agent Dialogue" (CRAG). We would like to thank Michael White and Scott Nowson for their assistance and our anonymous reviewers for their helpful comments.

References

Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James W. Pennebaker. 2005. Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*.

Gene Ball and Jack Breese. 2000. Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 189–219. MIT Press, Cambridge, MA, USA.

John A. Bateman and Cécile L. Paris. 2005. Adaptation to affective factors: architectural impacts for natural language generation and dialogue. In *Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*, Edinburgh, UK.

Allan Bell. 1984. Language style as audience design. *Language in Society*, 13(2):145–204.

Anja Belz. 2003. And now with feeling: Developments in emotional language generation. Technical Report ITRI-03-21, Information Technology Research Institute, University of Brighton, Brighton.

Susan E. Brennan. 1996. Lexical entrainment in spontaneous dialog. In *Proceedings of the 1996 International Symposium on Spoken Dialogue (ISSD-96)*, pages 41–44, Philadelphia, PA.

Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. 2005. 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 (UM-05)*, Edinburgh, UK.

Paul T. Costa and Robert R. McCrae, 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Jean-Marc Dewaele and Adrian Furnham. 1999. Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49:509–544.

Mary Ellen Foster and Michael White. 2004. Techniques for Text Planning with XSLT. In *Proc. of the 4th NLPXML Workshop*.

Simon Garrod and Gwyneth Doherty. 1994. Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3):181–215.

Patrick Gebhard. 2005. Alma: a layered model of affect. In *AAMAS '05: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 29–36, New York, NY, USA. ACM Press.

Alastair J. Gill, Annabel J. Harrison, and Jon Oberlander. 2004. Interpersonality: Individual differences and interpersonal priming. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 464–469.

Eduard Hovy. 1990. Pragmatics and natural language generation. *Artificial Intelligence*, 43.

Amy Isard, Carsten Brockmann, and Jon Oberlander. 2005. Re-creating dialogues from a corpus. In *Proceedings of the Workshop on Using Corpora for Natural Language Generation at Corpus Linguistics 2005 (CL-05)*, pages 7–12, Birmingham, UK.

Adam S. Kahn. 2006. Master's thesis, Stanford University.

A. Bryan Loyall and Joseph Bates. 1997. Personality-rich believable agents that use language. In J. Lewis and B. Hayes-Roth, editors, *Proceedings of the 1st International Conference on Autonomous Agents (Agents'97)*. ACM Press.

Gerald Matthews, Ian J. Deary, and Martha C. Whiteman. 2003. *Personality Traits*. Cambridge University Press, Cambridge, 2nd edition.

Youngme Moon and Clifford Nass. 1996. How "real" are computer personalities? *Communication Research*, 23:651–674.

Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings of CHI 2000, The Hague, Amsterdam, 2000*, pages 329–336.

Scott Nowson. 2006. *The Language of Weblogs: A study of genre and individual differences*. Ph.D. thesis, University of Edinburgh.

Jon Oberlander and Scott Nowson. 2006. Whose thumb is it anyway? Classifying author personality from weblog text. In *Proceedings of COLING/ACL-06: 44th Annual Meeting of the Association for Computational Linguistics and 21st International Conference on Computational Linguistics*, Sydney.

Daniel S. Paiva and Roger Evans. 2004. A framework for stylistically controlled generation. In *Proceedings of the 3rd International Conference on Natural Language Generation*, pages 120–129.

James W. Pennebaker and Laura King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77:1296–1312.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. *Linguistic Inquiry and Word Count 2001*. Lawrence Erlbaum Associates, Mahwah, NJ.

James W. Pennebaker, Matthias R. Mehl, and Kate G. Neiderhoffer. 2002. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54:547–577.

Martin J. Pickering and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4):633–651.

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

Paul Piwek and Kees van Deemter. 2003. Dialogue as discourse: Controlling global properties of scripted dialogue. In *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*.

Ehud Reiter and Somayajulu Sripada. 2004. Contextual influences on near-synonym choice. In *Proceedings of the Third International Conference on Natural Language Generation*, pages 161–170.

Thomas Rist, Elisabeth André, and Stephan Baldes. 2003. A flexible platform for building applications with life-like characters. In *IUI '03: Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 158–165, New York, NY, USA. ACM Press.

Klaus Scherer. 1979. Personality markers in speech. In K. R. Scherer and H. Giles, editors, *Social Markers in Speech*, pages 147–209. Cambridge University Press, Cambridge.

Andreas Stolcke, Harry Bratt, John Butzberger, Horacio Franco, Venkata Ramana Rao Gadde, Madelaine Plauché, Colleen Richey, Elizabeth Shriberg, Kemal Sönmez, Fuliang Weng, and Jing Zheng. 2000. The SRI March 2000 Hub-5 conversational speech transcription system. In *Proceedings of the 2000 Speech Transcription Workshop*, College Park, MD.

Andreas Stolcke. 2002. SRILM – an extensible language modeling toolkit. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP-02)*, pages 901–904, Denver, CO.

Sebastian Varges. 2005. Spatial descriptions as referring expressions in the MapTask domain. In *Proceedings of the 10th European Workshop on Natural Language Generation*.

Marilyn A. Walker, Janet E. Cahn, and Steve J. Whittaker. 1997. Improvising linguistic style: Social and affective bases for agent personality. In J. Lewis and B. Hayes-Roth, editors, *Proceedings of the 1st International Conference on Autonomous Agents (Agents'97)*, pages 96–105. ACM Press.

Michael White. 2004. Reining in CCG Chart Realization. In *Proceedings of the 3rd International Conference on Natural Language Generation*, pages 182–191.

Michael White. 2006. Efficient Realization of Coordinate Structures in Combinatory Categorial Grammar. *Research on Language & Computation*, online first, March.