Modelling alignment for affective dialogue

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
ABSTRACT

For a successful and satisfying interaction, a dialogue participant may align their language to be more like that of their interlocutor. In the first part of this paper, we examine the alignment phenomenon from the viewpoint of personality-related, linguistic, sociolinguistic and psycholinguistic research, concluding that some people are stronger aligners than others.

Motivated by these results, we describe an approach to modelling alignment behaviour in a natural language generation system, using the OpenCCG surface realiser [30], which allows utterance candidates to be ranked by n-gram language models. We investigate the extent to which alignment can be simulated using word sequences alone (not syntactic structures). To this end, we interpolate a default language model with one calculated on the basis of a cached sentence. Experiments on sentences with the prepositional/double object alternation show that varying the weight given to the cache model varies the propensity to align.

1. INTRODUCTION

A dialogue participant may – or may not – align their language to be more like that of their interlocutor. There is evidence that alignment (or the lack of it) has an impact on people's perceptions of the success of the dialogue. In general, people prefer interlocutors who align. Hence, alignment has affective consequences. In addition, there is preliminary evidence that stable affective properties of a dialogue participant (in particular, their level of Neuroticism) have an impact on the likelihood of their aligning with their interlocutor.

We are developing a natural language dialogue system whose alignment behaviour is parameterisable, using the OpenCCG surface realiser [30], a new open source realiser for Steedman’s Combinatory Categorial Grammar (CCG; [27]). The language models used by the realiser can be specialised to reflect language known to be associated with particular personality types. Together, variable alignment and specialised language models should help project coherent linguistic personalities.

Dialogue agents with linguistic personality have at least two main uses: an agent can interact directly with a human user; or two agents can interact with one another, with the human user as a third-party overseer. In the human-computer case, it seems plausible to suppose that it will always be better to build an agent that aligns to its user; Reeves, Nass and colleagues have argued that this leads to systems being rated as more helpful and more intelligent [22, 4]. But in the computer-computer case, the issue is less clear-cut. If part of the point of the interaction is to entertain the user, then an aspect of the entertainment may well lie in the distinctness of the personalities projected; and it is plausible to suppose that the tendency to acknowledge – or ignore – one’s partner’s way of speaking will make an impression on overseers, as has been found in observation of human-human dialogue [3].

The convincingness of the dialogue personalities has yet to be evaluated. However, we have carried out experiments on variable alignment, to establish which parameter settings will maximise a dialogue system’s alignment with its interlocutor – and hence, how to vary these settings to simulate different affective stances. The rest of this paper puts these experiments in context, via the following structure. First, a brief introduction to work on personality and language is given. Then, we discuss sociolinguistic and psycholinguistic perspectives on alignment, noting in particular recent work which suggests that propensity to align may depend upon personality. We then sketch our current approach to modelling alignment in OpenCCG realisation by exploiting simple n-gram language models. The results of our recent computational experiments are given, and we finish by discussing possible next steps.

2. PERSONALITY, DIALOGUE AND INTERPERSONAL PRIMING

2.1 Personality and Language

Individuals differ systematically in their communicative style. Some differences can be attributed to stable affective properties, such as personality traits, like Extraversion and Neuroticism (or Emotional Stability). These are the two dimensions of variation which are common to the leading trait theories of personality: Costa and McCrae’s five factor model [6]; and Eysenck’s three factor model [9]. Level of Extraversion is associated with degree of outgoingness and assertiveness; level of Neuroticism is associated with degree of anxiousness, and self-consciousness [21].
These personality traits have been found to be associated with significant differences between individuals’ language production behaviours [24, 8]. According to Dewaele and collaborators, a technical notion of formality is associated with preference for, for instance, nouns and adjectives, as opposed to verbs and adverbs; and preference for formality is related to level of Extraversion [8, 18]. Recent data-driven work has investigated text in e-mail and weblogs, and suggested that there are word n-grams, and parts-of-speech n-grams associated with each end (High or Low) of both dimensions (Extravert or Neurotic) [13, 14]. In addition, other personality dimensions have also been found to make a difference. For instance, Nowson and colleagues report that, in weblogs, degree of Agreeableness (or affectionateness) appears to bear a stronger association (than Extraversion or Neuroticism) with the ‘formality’ of a speaker’s language [23]. Arguably, more Agreeable weblog writers take account of the differing contexts of writer and reader, and hence use more formal, less contextual language. This result, in asynchronous communication, naturally raises the issue of the extent to which synchronous communicators take account of their interlocutor’s presence. It is to this more specific topic that we now turn.

2.2 Dialogue and Priming

Dialogue participants orient towards their interlocutors at a number of levels [26]. The phenomena have both social and cognitive facets. On the social side, a key focus of interest is co-operation and audience design. On the cognitive side, a key focus is co-ordination and interpersonal priming.

Sociolinguistic studies show that speakers adapt their phonological, lexical, and syntactic choices to ones appropriate to their intended audience [20, 7, 1, 2]. Such audience design is considered to be at least partially under conscious control, reflecting a speaker’s co-operation or willingness to adopt another’s perspective [17]. Co-operative speakers are seen by third parties to be better communicators [3]. By contrast, cognitive scientists tend to see alignment or co-ordination as an artifact of underlying, unconscious, language production processes [10, 11]. Co-ordination can be probed via experiments on interpersonal priming. By using a confederate methodology, we can establish to what extent subjects are more likely to use a particular item after they have heard their interlocutor use that item. The item might be a word or a particular grammatical construction. Focusing on the latter, experiments have investigated passives, and ditransitives [25].

Recent work suggests that some people are stronger aligners than others [12]. In particular, the study investigated subjects’ propensity to use a passive construction after hearing their interlocutor use one. It was expected that Extravert sociability would be related to the strength of priming effects, although it was also proposed that high levels of Neuroticism would suppress priming. In the event, results indicated that Extraversion has no effect, but Neuroticism does have an effect. However, a non-linear effect emerged: it transpired that both High and Low levels of Neuroticism led to weaker priming. It was Mid levels that led to significantly stronger priming; see Figure 1.

So, the question arises: can we build a natural language generation system that can be parameterised so as to simulate different types of aligners?

3. MODELLING ALIGNMENT

3.1 Natural Language Generation Guided by N-Gram Language Models

Our goal is to investigate whether the effect of syntactic priming can be simulated by looking at word sequences alone,
and to integrate such alignment behaviour into a natural language generation system.

The OpenCCG surface realiser takes as input a logical form specifying the propositional meaning of a sentence, and returns a ranked list of surface strings that express this meaning according to the lexicon and grammar, where rankings are determined by n-gram language models derived from examples of desired realisations (reducing the need for handcrafted rules). The n-gram models are employed in a best-first anytime search, in such a way that preferred realisations tend to be found early in the search process. It is possible to plug in n-gram models that interpolate a cache, with the effect that aligning realisations will be produced with less effort.

Cache models work by interpolating simple language models derived from the recent context with more elaborate, context-independent models. As Goodman [16] explains, cache models can yield impressive reductions in perplexity, and bigram and trigram cache models usually work better than unigram ones.

We use the SRILM toolkit [28] to compute n-gram language models and score test sentences. The toolkit includes a trigram language model derived from Switchboard [15] corpus data, which we use as a default smoothed language model. To simulate alignment, we interoperate it with a language model calculated on the basis of a cached sentence. The cached sentence can be seen as the previous utterance in a dialogue.

3.2 Interpolating Cache and Default Language Models

By the chain rule, the probability of a word sequence \( w_1, \ldots, w_n \), \( w_n \) is equal to the product of the probabilities of each word \( w_i \) given the preceding ones, \( w_i^{-1} \):

\[
P(w_1, \ldots, w_n) = \prod_{i=1}^{n} P(w_i | w_i^{-1})
\]

We approximate the probability of a word given its history by the probability of a word given the preceding two words, i.e. by using trigrams. In the trigram probability, we interoperate a cache model, \( P_{\text{cache}}(w_i | w_i^{-2}) \), with the default smoothed model, \( P_{\text{smooth}}(w_i | w_i^{-2}) \), as follows:

\[
P(w_i | w_i^{-1}) \approx \lambda P_{\text{cache}}(w_i | w_i^{-2}) + (1 - \lambda) P_{\text{smooth}}(w_i | w_i^{-2})
\]

The cache model gets weight \( \lambda \), and the default model 1 – \( \lambda \). The cache model itself is the uniform interpolation of word- and class-based trigram models, as shown in (3) below. Classes are a way of grouping together lexical items with similar semantic properties; they provide a backoff mechanism if there is no exact word match.

\[
P_{\text{cache}}(w_i | w_i^{-2}) = \frac{P_{\text{word}}(w_i | w_i^{-2}) + P_{\text{class}}(c_i | w_i^{-2}) P(w_i | c_i)}{2}
\]

The word- and class-based trigram models are themselves uniform interpolations of unigram, bigram and trigram probabilities:

\[
P_{\text{word}}(w_i | w_i^{-2}) = \frac{P(w_i) + P(w_i | w_i^{-1}) + P(w_i | w_i^{-2})}{3}
\]

\[
P_{\text{class}}(c_i | w_i^{-2}) = \frac{P(c_i) + P_{\text{class}}(c_i | w_i^{-1}) + P_{\text{class}}(c_i | w_i^{-2})}{3}
\]

With the class-based bigrams and trigrams, the probability of the current class \( c_i \) given the previous word(s) is backed off (again via uniform interpolation) to the probability given the previous class(es):

\[
P_{\text{class}}(c_i | w_i^{-1}) = \frac{P(c_i | w_i^{-1}) + P(c_i | c_i^{-1})}{2}
\]

With this model, varying \( \lambda \) varies the propensity to align.

4. EXPERIMENT: PO/DO PRIMING

4.1 Psycholinguistic Evidence

A typical syntactic priming construction studied in the psycholinguistic literature is the prepositional object (PO) / double object (DO) alternation. Branigan et al. [5] found that the syntactic structure used by a confederate to describe a picture influenced the syntactic structure of the experimental subject’s subsequent picture description. After hearing a PO prime, they were more likely to also use a PO construction, while DO primes elicited DO target descriptions. The priming effect was stronger when the verb for prime and target picture description remained the same.

4.2 Materials

Using the verbs and noun phrases of a sentence completion experiment on syntactic priming [25] as a dictionary, we randomly generated combinations of a prime and four target sentences. An example PO prime is (8):

(8) the manager loaned the towel to the kidnapper

Target sentences (9a) and (9c) also exhibit the PO structure, while (9b) and (9d) are DO constructions. Targets (9a) and (9b) keep the same verb, while a different verb is chosen for (9c) and (9d).

(9a) the sailor loaned the book to the professor
(9b) the sailor loaned the professor the book
(9c) the sailor gave the book to the professor
(9d) the sailor gave the professor the book

Each word was assigned a semantic class. All verbs had the class exchange_verb; nouns were either animate or inanimate, and function words each had a class of their own (the or to).

4.3 Design

The prime sentence was used to train language models that were combined into a cache model and then interpolated with the default model, as specified above. We varied the weight \( \lambda \) from 0.1 to 0.9 in steps of 0.1 and observed the log probabilities assigned to the target sentences for each setting.
We observe that the interpolated language model prefers more probable. As predicted, modifying the weight given to the cache model by changing the $\lambda$ value influences alignment behaviour. The log probabilities of target sentences (9a)–(9d) given the cached PO sentence (8), for varying $\lambda$ values. For some stimuli, there is a switch as early as $\lambda = 0$, and some do not switch over at all. In general, though, the probabilities always move towards alignment when $\lambda$ is increased.

Furthermore, for each $\lambda$, the sentences with the same verb as the prime always get rated higher than those with the different verb.

Looking at other examples with PO primes, we note that the switch from non-alignment to alignment occurs at different $\lambda$ values. For some stimuli, there is a switch as early as $\lambda = 0.2$, for others only at $\lambda = 0.8$, and some do not switch over at all. In general, though, the probabilities always move towards alignment when $\lambda$ is increased.

The default language model generally seems to prefer DO sentences. Therefore, in conditions using a DO prime, we get alignment throughout; the PO targets are never preferred over the DO ones, regardless of the $\lambda$ setting – but still, as $\lambda$ is being increased, the difference between the DO and PO target probabilities increases, which signifies a stronger alignment effect.

Table 1: Log probabilities of target sentences (9a)–(9d) given the cached PO sentence (8), for varying $\lambda$ values. For some stimuli, there is a switch as early as $\lambda = 0$, and some do not switch over at all. In general, though, the probabilities always move towards alignment when $\lambda$ is increased.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>same verb prep. obj.</th>
<th>same verb double obj.</th>
<th>diff. verb prep. obj.</th>
<th>diff. verb double obj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>-20.27</td>
<td>-19.81</td>
<td>-22.96</td>
<td>-22.51</td>
</tr>
<tr>
<td>0.2</td>
<td>-19.32</td>
<td>-19.11</td>
<td>-22.40</td>
<td>-22.20</td>
</tr>
<tr>
<td>0.3</td>
<td>-18.80</td>
<td>-18.76</td>
<td>-22.14</td>
<td>-22.10</td>
</tr>
<tr>
<td>0.4</td>
<td>-18.51</td>
<td>-18.59</td>
<td>-22.05</td>
<td>-22.13</td>
</tr>
<tr>
<td>0.5</td>
<td>-18.37</td>
<td>-18.56</td>
<td>-22.09</td>
<td>-22.28</td>
</tr>
<tr>
<td>0.6</td>
<td>-18.36</td>
<td>-18.64</td>
<td>-22.26</td>
<td>-22.55</td>
</tr>
<tr>
<td>0.7</td>
<td>-18.50</td>
<td>-18.87</td>
<td>-22.60</td>
<td>-22.97</td>
</tr>
<tr>
<td>0.8</td>
<td>-18.85</td>
<td>-19.31</td>
<td>-23.19</td>
<td>-23.65</td>
</tr>
<tr>
<td>0.9</td>
<td>-19.02</td>
<td>-20.21</td>
<td>-24.32</td>
<td>-24.90</td>
</tr>
</tbody>
</table>

4.4 Results
As predicted, modifying the weight given to the cache model by changing the $\lambda$ value influences alignment behaviour.

How would the natural language generation system decide whether to align or not to align? Given the probabilities' general trend towards alignment for higher $\lambda$ values, we suggest to employ a strategy of sampling from low-cost utterances, as described by Stone et al. [29]. Instead of always choosing the single utterance with the lowest cost (i.e. highest probability), the idea is to perturb the costs by small random amounts. This way, we would get more or less frequent alignment, weighted by the probabilities, instead of an all or nothing behaviour.

5. CONCLUSION AND FUTURE WORK
Arguably, for a co-operative, satisfying interaction, it is always preferable to align than not to do so. However, differing people may align to greater or lesser extents. To simulate that range of behaviour, we developed an approach to combining a default language model with a cache model trained on recent dialogue context. The alignment effect is achieved by varying the weight given to the cache model and by sampling from low-cost utterances.

We plan to experiment with cache models of richer part-of-speech tags, or supertags [19], to see if they give further improvement. With CCG (and related lexicalised theories of syntax), supertags encode the syntactic category of a lexical item. This information provides an extended domain of locality, and the model would be less dependent on exact word sequences.

Of course, alignment is only one aspect of an individual’s dialogue behaviour. Before they align (or fail to do so), an individual has distinctive language behaviour. We can capture these differences by training OpenCCG’s language models on language from particular personality types. A strong aligner will give more emphasis to conversational context (including their interlocutor’s language); a weak aligner will stick more closely to their own language model. We intend shortly to carry out user evaluation of variable alignment both with, and without, such personality-based language models.

6. ACKNOWLEDGEMENTS
This research has been funded by Scottish Enterprise through the Edinburgh-Stanford Link project “Critical Agent Dialogue” (CrAg). Our special thanks to Alastair Gill and Annabel Harrison.

7. REFERENCES


