Modelling alignment for affective dialogue

Modelling Alignment for Affective Dialogue

Carsten Brockmann
Carsten.Brockmann@ed.ac.uk

Amy Isard
Amy.Isard@ed.ac.uk

Jon Oberlander
J.Oberlander@ed.ac.uk

Michael White
Michael.White@ed.ac.uk

School of Informatics, University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW, UK

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].

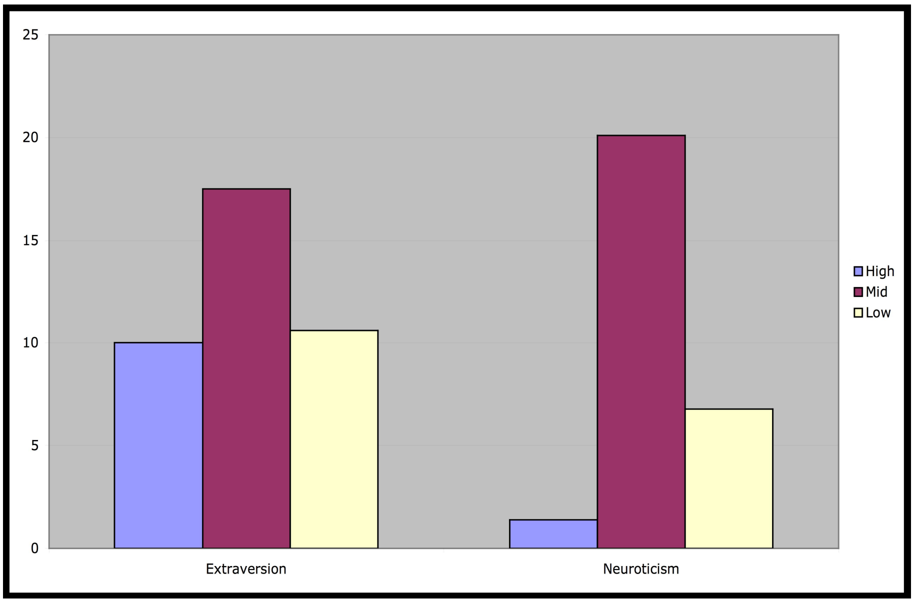

Figure 1: From Gill et al. (2004): Percentage priming by personality group. Differences between Extravert groups are not significant at $p < 0.05$; differences between Mid-Neurotics and other Neurotic groups are significant at $p < 0.05$.

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 Neurotic emotionality might 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 interpolate 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 is equal to the product of the probabilities of each word w_i given the preceding ones, w_1^{i-1}:

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

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 interpolate a cache model, $P_{\text{cache}}(w_i|w_{i-2}^{i-1})$, with the default smoothed model, $P_{\text{smooth}}(w_i|w_{i-2}^{i-1})$, as follows:

(2) $$P(w_i|w_1^{i-1}) \approx \lambda P_{\text{cache}}(w_i|w_{i-2}^{i-1}) + (1-\lambda) P_{\text{smooth}}(w_i|w_{i-2}^{i-1})$$

The cache model gets weight λ, 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.

(3) $$P_{\text{cache}}(w_i|w_{i-2}^{i-1}) = \frac{P_{\text{word}}(w_i|w_{i-2}^{i-1}) + P_{\text{class}}(c_i|w_{i-2}^{i-1}) P(w_i|c_i)}{2}$$

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

(4) $$P_{\text{word}}(w_i|w_{i-2}^{i-1}) = \frac{P(w_i) + P(w_i|w_{i-1}) + P(w_i|w_{i-2}^{i-1})}{3}$$

(5) $$P_{\text{class}}(c_i|w_{i-2}^{i-1}) = \frac{P(c_i) + P_{\text{bi}}(c_i|w_{i-1}) + P_{\text{tri}}(c_i|w_{i-2}^{i-1})}{3}$$

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):

(6) $$P_{\text{bi}}(c_i|w_{i-1}) = \frac{P(c_i|w_{i-1}) + P(c_i|c_{i-1})}{2}$$

(7) $$P_{\text{tri}}(c_i|w_{i-2}^{i-1}) = \frac{P(c_i|w_{i-2}^{i-1}) + P(c_i|w_{i-2}, c_{i-1}) + P(c_i|c_{i-2}^{i-1})}{3}$$

With this model, varying λ 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 λ from 0.1 to 0.9 in steps of 0.1 and observed the log probabilities assigned to the target sentences for each setting.

Table 1: Log probabilities of target sentences (9a)–(9d) given the cached PO sentence (8), for varying λ

λ	same verb prep. obj.	same verb double obj.	diff. verb prep. obj.	diff. verb double obj.
0.1	−20.27	**−19.81**	−22.96	**−22.51**
0.2	−19.32	**−19.11**	−22.40	**−22.20**
0.3	−18.80	**−18.76**	−22.14	**−22.10**
0.4	**−18.51**	−18.59	**−22.05**	−22.13
0.5	**−18.37**	−18.56	**−22.09**	−22.28
0.6	**−18.36**	−18.64	**−22.26**	−22.55
0.7	**−18.50**	−18.87	**−22.60**	−22.97
0.8	**−18.85**	−19.31	**−23.19**	−23.65
0.9	**−19.62**	−20.21	**−24.32**	−24.90

4.4 Results

As predicted, modifying the weight given to the cache model by changing the λ value influences alignment behaviour. The log probabilities of target sentences (9a)–(9d) given the cached PO sentence (8), for different values of λ, are shown in Table 1; a value closer to zero means that a sentence is more probable.

We observe that the interpolated language model prefers the sentences with the same syntactic structure (aligns) for λ values of 0.4 and above, while there is no alignment for lower λ values. The preferred choice for each value of λ is highlighted.

Furthermore, for each λ, 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 λ 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 λ 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 λ setting – but still, as λ is being increased, the difference between the DO and PO target probabilities increases, which signifies a stronger alignment effect.

4.5 Discussion

The findings show that alignment behaviour can be simulated by interpolating a cache model trained on recently used language with a default language model and varying the respective models' weights.

We also observe that target utterances with the same verb as the prime are preferred over utterances with a different verb. This mirrors human performance in psycholinguistic experiments, where it was found that alignment was more likely to occur in same-verb conditions.

The moving λ values for the switch-over point from non-alignment to alignment may be explained by considerably varying prior probabilities of the respective word sequences in the default language model. On the other hand, it could be that the Switchboard model we are using is not very representative of some of these words.

How would the natural language generation system decide whether to align or not to align? Given the probabilities' general trend towards alignment for higher λ 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

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

[2] J. Bradac and R. Wisegarver. Ascribed status lexical diversity, and accent: Determinants of perceived status, solidarity, and control of speech style. *Journal of Language and Social Psychology*, 3:239–255, 1984.

[3] J. J. Bradac. Language attitudes and impression formation. In H. Giles and W. P. Robinson, editors,

Handbook of Language and Social Psychology, chapter 19, pages 387–412. John Wiley & Sons, Chichester, UK, 1990.

[4] H. Branigan, M. Pickering, J. Pearson, J. McLean, and C. Nass. 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*, pages 186–191, 2003.

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

[6] P. Costa and R. R. McCrae. *NEO PI-R Professional Manual*. Psychological Assessment Resources, Odessa, Florida, 1992.

[7] N. Coupland. Style-shifting in a Cardiff work-setting. *Language in Society*, 9:1–12, 1980.

[8] J.-M. Dewaele and A. Furnham. Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49:509–544, 1999.

[9] H. Eysenck and S. B. G. Eysenck. *The Eysenck Personality Questionnaire-Revised*. Hodder and Stoughton, Sevenoaks, 1991.

[10] S. Garrod and G. Doherty. Saying what you mean in dialogue. *Cognition*, 27:181–218, 1987.

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

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

[13] A. Gill and J. Oberlander. Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 363–368, 2002.

[14] A. Gill and J. Oberlander. Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; Neuroticism is a worry. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, pages 456–461, 2003.

[15] J. J. Godfrey, E. C. Holliman, and J. McDaniel. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP-92)*, pages 517–520, San Francisco, CA, 1992.

[16] J. T. Goodman. A bit of progress in language modeling. *Computer Speech and Language*, 15(4):403–434, 2001.

[17] S. Haywood, M. Pickering, and H. Branigan. Co-operation and co-ordination in the production of noun phrases. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, pages 533–538, 2003.

[18] F. Heylighen and J.-M. Dewaele. Variation in the contextuality of language: An empirical measure. *Foundations of Science*, 7:293–340, 2002.

[19] A. K. Joshi and B. Srinivas. Disambiguation of super parts of speech (or supertags): Almost parsing. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, pages 154–160, Kyoto, Japan, 1994.

[20] W. Labov. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia, 1972.

[21] G. Matthews, I. Deary, and M. Whiteman. *Personality Traits*. Cambridge University Press, Cambridge, 2nd edition, 2003.

[22] C. Nass, Y. Moon, B. Fogg, and B. Reeves. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43:223–239, 1995.

[23] S. Nowson, J. Oberlander, and A. Gill. Weblogs, genres and individual differences. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, Stresa, Italy, 2005.

[24] J. W. Pennebaker and L. King. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77:1296–1312, 1999.

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

[26] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–225, 2004.

[27] M. Steedman. *The Syntactic Process*. Language, Speech, and Communication. MIT Press, Cambridge, MA, 2000.

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

[29] M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees, and C. Bregler. Speaking with hands: Creating animated conversational characters from recordings of human performance. *ACM Transactions on Graphics*, 23(3):506–513, 2004.

[30] M. White. Efficient realization of coordinate structures in combinatory categorial grammar. *Research on Language and Computation*, 2004. To appear.