

Perception of Personality and Naturalness through Dialogues by Native Speakers of American English and Arabic

Maxim Makatchev
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
mmakatch@cs.cmu.edu

Reid Simmons
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
reids@cs.cmu.edu

Abstract

Linguistic markers of personality traits have been studied extensively, but few cross-cultural studies exist. In this paper, we evaluate how native speakers of American English and Arabic perceive personality traits and naturalness of English utterances that vary along the dimensions of verbosity, hedging, lexical and syntactic alignment, and formality. The utterances are the turns within dialogue fragments that are presented as text transcripts to the workers of Amazon’s Mechanical Turk. The results of the study suggest that all four dimensions can be used as linguistic markers of all personality traits by both language communities. A further comparative analysis shows cross-cultural differences for some combinations of measures of personality traits and naturalness, the dimensions of linguistic variability and dialogue acts.

1 Introduction

English has been used as a lingua franca across the world, but the usage differs. The variabilities in English introduced by dialects, cultures, and non-native speakers result in different syntax and words expressing similar meanings and in different meanings attributed to similar expressions. These differences are a source of *pragmatic failures* (Thomas, 1983): situations when listeners perceive meanings and affective attitudes unintended by speakers. For example, Thomas (1984) reports that usage of Illocutionary Force Indicating Devices (IFIDs, such as “I warn you”, (Searle, 1969)) in English by native speakers of Russian causes the speakers to sometimes

appear “inappropriately domineering in interactions with English-speaking equals.” Dialogue systems, just like humans, may misattribute attitudes and misinterpret intent of user’s utterances. Conversely, they may also cause misattributions and misinterpretations on the user’s part. Hence, taking into account the user’s dialect, culture, or native language may help reduce pragmatic failures.

This kind of adaptation requires a mapping from utterances, or more generally, their linguistic features, to meanings and affective attributions for each of the target language communities. In this paper we present an exploratory study that evaluates such a mapping from the linguistic features of verbosity, hedging, alignment, and formality (as defined in Section 3.1) to the perceived personality traits and naturalness across the populations of native speakers of American English and Arabic.

Estimating the relationship between linguistic features and their perception across language communities faces a number of methodological difficulties. First, language communities shall be outlined, in a way that will afford generalizing within their populations. Defining language communities is a hard problem, even if it is based on the “mother tongue” (McPherson et al., 2000). Next, linguistic features that are potentially important for the adaptation must be selected. These are, for example, the linguistic devices that contribute to realization of *rich points* (Agar, 1994), i.e. the behaviors that signal differences between language communities. To be useful for dialogue system research, the selected linguistic features should be feasible to implement in natural language generation and interpretation mod-

ules. Then, a corpus of stimuli that span the variability of the linguistic features must be created. The stimuli should reflect the context where the dialogue system is intended to be used. For example, in case of an information-giving dialogue system, the stimuli should include some question-answer adjacency pairs (Schegloff and Sacks, 1973). Finally, scales should be chosen to allow for scoring of the stimuli with respect to the metrics of interest. These scales should be robust to be applied within each of the language communities.

In the remainder of this paper, we describe each of these steps in the context of an exploratory study that evaluates perception of English utterances by native speakers of American English and Arabic. Our application is an information-giving dialogue system that is used by the robot receptionists (roboceptionists) in Qatar and the United States (Makatchev et al., 2009; Makatchev et al., 2010). In the next section, we continue with an overview of the related work. Section 3 introduces the experiment, including the selection of stimuli, measures, design, and describes the recruitment of participants via Amazon's Mechanical Turk (MTurk). We discuss results in Section 4 and provide a conclusion in Section 5.

2 Related work

2.1 Cross-cultural variability in English

Language is tightly connected with culture (Agar, 1994). As a result, even native speakers of a language use it differently across dialects (e.g. African American Vernacular English and Standard American English), genders (see, for example, (Lakoff, 1973)) and social statuses (e.g. (Huspek, 1989)), among other dimensions.

Speakers of English as a second language display variabilities in language use that are consistent with their native languages and backgrounds. For example, Nelson et al. (1996) reports that Syrian speakers of Arabic tend to use different compliment response strategies as compared with Americans. Aguilar (1998) reviews types of pragmatic failures that are influenced by native language and culture. In particular, he cites Davies (1987) on a pragmatic failure due to *non-equivalence of formulas*: native speakers of Moroccan Arabic use a spoken formulaic expression to wish a sick person quick recovery, whereas in

English the formula “get well soon” is not generally used in speech. Feghali (1997) reviews features of Arabic communicative style, including indirectness (concealment of wants, needs or goals (Gudykunst and Ting-Toomey, 1988)), elaborateness (rich and expressive language use, e.g. involving rhetorical patterns of exaggeration and assertion (Patai, 1983)) and affectiveness (i.e. “intuitive-affective style of emotional appeal” (Glenn et al., 1977), related to the patterns of organization and presentation of arguments).

In this paper, we are concerned with English usage by native speakers of American English and native speakers of Arabic. We have used the features of the Arabic communicative style outlined above as a guide in selecting the dimensions of linguistic variability that are presented in Section 3.1.

2.2 Measuring pragmatic variation

Perception of pragmatic variation of spoken language and text has been shown to vary across cultures along the dimensions of personality (e.g. (Scherer, 1972)), emotion (e.g. (Burkhardt et al., 2006)), deception (e.g. (Bond et al., 1990)), among others. Within a culture, personality traits such as extraversion, have been shown to have consistent markers in language (see overview in (Mairesse et al., 2007)). For example, Furnham (1990) notes that in conversation, extraverts are less formal and use more verbs, adverbs and pronouns. However, the authors are not aware of any quantitative studies that compare linguistic markers of personality across cultures. The present study aims to help fill this gap.

A mapping between linguistic dimensions and personality has been evaluated by grading essays and conversation extracts (Mairesse et al., 2007), and by grading utterances generated automatically with a random setting of linguistic parameters (Mairesse and Walker, 2008). In the exploratory study presented in this paper, we ask our participants to grade dialogue fragments that were manually created to vary along each of the four linguistic dimensions (see Section 3.1).

3 Experiment

In the review of related work, we presented some evidence supporting the claim that linguistic markers of personality may differ across cultures. In this section, we describe a study that evaluates perception of personality traits and naturalness of utterances by native speakers of American English and Arabic.

3.1 Stimuli

The selection of stimuli attempts to satisfy three objectives. First, our application: our dialogue system is intended to be used on a robot receptionist. Hence, the stimuli are snippets of dialogue that include four dialogue acts that are typical in this kind of embodied information-giving dialogue (Makatchev et al., 2009): a greeting, a question-answer pair, a disagreement (with the user’s guess of an answer), and an apology (for the robot not knowing the answer to the question).

Second, we would like to vary our stimuli along the linguistic dimensions that are potentially strong indicators of personality traits. Extraverts, for example, are reported to be more verbose (use more words per utterances and more dialogue turns to achieve the same communicative goal), less formal (Furnham, 1990) (in choice of address terms, for example), and less likely to hedge (use expressions such as “perhaps” and “maybe”) (Nass et al., 1995). Lexical and syntactic alignment, namely, the tendency of a speaker to use the same lexical and syntactic choices as their interlocutor, is considered, at least in part, to reflect the speaker’s co-operation and willingness to adopt the interlocutor’s perspective (Haywood et al., 2003). There is some evidence that the degree of alignment is associated with personality traits of the speakers (Gill et al., 2004).

Third, we would like to select linguistic dimensions that potentially expose cross-cultural differences in perception of personality and naturalness. In particular, we are interested in the linguistic devices that help realize *rich points* (the behaviors that signal differences) between the native speakers of American English and Arabic. We choose to realize indirectness and elaborateness, characteristic of Arabic spoken language (Feghali, 1997), by varying the dimensions of verbosity and hedging. High *power distance*, or influence of relative social status

on the language (Feghali, 1997), can be realized by the degrees of formality and alignment.

In summary, the stimuli are dialogue fragments where utterances of one of the interlocutors vary across (1) dialogue acts: a greeting, question-answer pair, disagreement, apology, and (2) four linguistic dimensions: verbosity, hedging, alignment, and formality. Each of the linguistic dimensions is parameterized by 3 values of valence: negative, neutral and positive. Within each of the four dialogue acts, stimuli corresponding to the neutral valences are represented by the same dialogue across all four linguistic dimensions. The four linguistic dimensions are realized as follows:

- Verbosity is realized as number of words within each turn of the dialogue. In the case of the greeting, positive verbosity is realized by increased number of dialogue turns.¹
- Positive valence of hedging implies more tentative words (“maybe,” “perhaps,” etc.) or expressions of uncertainty (“I think,” “if I am not mistaken”). Conversely, negative valence of hedging is realized via words “sure,” “definitely,” etc.
- Positive valence of alignment corresponds to preference towards the lexical and syntactic choices of the interlocutor. Conversely, negative alignment implies less overlap in lexical and syntactic choices between the interlocutors.
- Our model of formality deploys the following linguistic devices: in-group identity markers that target positive face (Brown and Levinson, 1987) such as address forms, jargon and slang, and deference markers that target negative face, such as “kindly”, terms of address, hedges. These devices are used in Arabic politeness phenomena (Farahat, 2009), and there is an evidence of their pragmatic transfer from Arabic to English (e.g. (Bardovi-Harlig et al., 2007) and (Ghawi, 1993)). The set of stimuli that vary along the formality are presented in Table 2.

Each dialogue fragment is presented as a text on

¹The multi-stage greeting dialogue was developed via ethnographic studies conducted at Alelo by Dr. Suzanne Wertheim. Used with permission from Alelo, Inc.

an individual web page. On each page, the participant is asked to imagine that he or she is one of the interlocutors and the other interlocutor is described as “a female receptionist in her early 20s and of the same ethnic background” as that of the participant. The description of the occupation, age, gender and ethnicity of the interlocutor whose utterances the participant is asked to evaluate should provide minimal context and help avoid variability due to the implicit assumptions that subjects may make.

3.2 Measures

In order to avoid a possible interference of scales, we ran two versions of the study in parallel. In one version, participants were asked to evaluate the receptionist’s utterances with respect to measures of the Big Five personality traits (John and Srivastava, 1999), namely the traits of extraversion, agreeableness, conscientiousness, emotional stability, and openness, using the ten-item personality questionnaire (TIPI, see (Gosling et al., 2003)). In the other version, participants were asked to evaluate the receptionist’s utterances with respect to their naturalness on a 7-point Likert scale by answering the question “Do you agree that the receptionist’s utterances were natural?” The variants of such a naturalness scale were used by Burkhardt et al. (2006) and Mairesse and Walker (2008).

3.3 Experimental design

The experiment used a crossed design with the following factors: dimensions of linguistic variability (verbosity, hedging, alignment, or formality), valence (negative, neutral, or positive), dialogue acts (greeting, question-answer, disagreement, or apology), native language (American English or Arabic) and gender (male or female).

In an attempt to balance the workload of the participants, depending on whether the participant was assigned to the study that used personality or naturalness scales, the experimental sessions consisted of one or two linguistic variability conditions—12 or 24 dialogues respectively. Hence valence and dialogue act were within-subject factors, while linguistic variability dimension were treated as an across-subject factor, as well as native language and gender. Within each session the items were presented in

Language	Country	<i>N</i>
Arabic	Algeria	1
	Bahrain	1
	Egypt	56
	Jordan	32
	Morocco	45
	Palestinian Territory	1
	Qatar	1
	Saudi Arabia	5
	United Arab Emirates	13
	Total	155
American English	United States	166

Table 1: Distribution of study participants by country.

a random order to minimize possible carryover effects.

3.4 Participants

We used Amazon’s Mechanical Turk (MTurk) to recruit native speakers of American English from the United States and native speakers of Arabic from any of the set of predominantly Arabic-speaking countries (according to the IP address).

Upon completion of each task, participants receive monetary reward as a credit to their MTurk account. Special measures were taken to prevent multiple participation of one person in the same study condition: the study website access would be refused for such a user based on the IP address, and MTurk logs were checked for repeated MTurk user names to detect logging into the same MTurk account from different IP addresses. Hidden questions were planted within the study to verify the fluency in the participant’s reported native language.

The distribution of the participants across countries is shown in Table 1. We observed a regional gender bias similar to the one reported by Ross et al. (2010): there were 100 male and 55 female participants in the Arabic condition, and 63 male and 103 female participants in the American English condition.

4 Results

We analyzed the data by fitting linear mixed-effects (LME) models (Pinheiro and Bates, 2000) and performing model selection using ANOVA. The comparison of models fitted to explain the personality

and naturalness scores (controlling for language and gender), shows significant main effects of valence and dialogue acts for all pairs of personality traits (and naturalness) and linguistic features. The results also show that for every personality trait (and naturalness) there is a linguistic feature that results in a significant three-way interaction between its valence, the native language, and the dialogue act. These results suggest that (a) for both language communities, every linguistic dimension is associated with every personality trait and naturalness, for at least some of the dialogue acts, (b) there are differences in the perception of every personality trait and naturalness between the two language communities.

To further explore the latter finding, we conducted a post-hoc analysis consisting of paired t-tests that were performed pairwise between the three values of valence for each combination of language, linguistic feature, and personality trait (and naturalness). Note, that comparing raw scores between the language conditions would be prone to find spurious differences due to potential culture-specific tendencies in scoring on the Likert scale: (a) perception of magnitudes and (b) appropriateness of the intensity of agreeing or disagreeing. Instead, we compare the language conditions with respect to (a) the relative order of the three valences and (b) the binarized scores, namely whether the score is above 4 or below 4 (with scores that are not significantly different from 4 excluded from comparison), where 4 is the neutral point of the 7-point Likert scale.

The selected results of the post-hoc analysis are shown in Figure 1. The most prominent cross-cultural differences were found in the scoring of naturalness across the valences of the formality dimension. Speakers of American English, unlike the speakers of Arabic, find formal utterances unnatural in greetings, question-answer and disagreement dialogue acts. Formal utterances tend to also be perceived as indicators of openness (omitted from the plot) and conscientiousness by Arabic speakers, and not by American English speakers, in disagreements and apologies respectively. Finally, hedging in apologies is perceived as an indicator of agreeableness by American English speakers, but not by speakers of Arabic.

Interestingly, no qualitative differences across language conditions were found in the perception

of extraversion and stability. It is possible that this cross-cultural consistency confirms the view of the extraversion, in particular, as one of most consistently identified dimensions (see, for example, (Gill and Oberlander, 2002)). It could also be possible that our stimuli were unable to pinpoint the extraversion-related rich points due to a choice of the linguistic dimensions or particular wording chosen. A larger variety of stimuli per condition, and an ethnography to identify potentially culture-specific linguistic devices of extraversion, could shed the light on this issue.

5 Conclusion

We presented an exploratory study to evaluate a set of linguistic markers of Big Five personality traits and naturalness across two language communities: native speakers of American English living in the US, and native speakers of Arabic living in one of the predominantly Arabic-speaking countries of North Africa and Middle East. The results suggest that the four dimensions of linguistic variability are recognized as markers of all five personality traits by both language communities. A comparison across language communities uncovered some qualitative differences in the perception of openness, conscientiousness, agreeableness, and naturalness.

The results of the study can be used to adapt natural language generation and interpretation to native speakers of American English or Arabic. This exploratory study also supports the feasibility of the crowdsourcing approach to validate the linguistic devices that realize rich points—behaviors that signal differences across languages and cultures.

Future work shall evaluate effects of regional dialects and address the issue of particular wording choices by using multiple stimuli per condition.

Acknowledgments

This publication was made possible by the support of an NPRP grant from the Qatar National Research Fund. The statements made herein are solely the responsibility of the authors.

The authors are grateful to Ameer Ayman Abdulsalam, Michael Agar, Hatem Alismail, Justine Caspell, Majd Sakr, Nik Melchior, and Candace Sidner for their comments on the study.

References

- Michael Agar. 1994. *Language shock: Understanding the culture of conversation*. William Morrow, New York.
- Maria Jose Coperias Aguilar. 1998. Intercultural (mis)communication: The influence of L1 and C1 on L2 and C2. A tentative approach to textbooks. *Cuadernos de Filología Inglesa*, 7(1):99–113.
- Kathleen Bardovi-Harlig, Marda Rose, and Edelmira L. Nickels. 2007. The use of conventional expressions of thanking, apologizing, and refusing. In *Proceedings of the 2007 Second Language Research Forum*, pages 113–130.
- Charles F. Bond, Adnan Omar, Adnan Mahmoud, and Richard Neal Bonser. 1990. Lie detection across cultures. *Journal of Nonverbal Behavior*, 14:189–204.
- P. Brown and S. C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press, Cambridge.
- F. Burkhardt, N. Audibert, L. Malatesta, O. Trk, Arslan, L., and V Auberge. 2006. Emotional prosody—does culture make a difference? In *Proc. Speech Prosody*.
- Eirlys E. Davies. 1987. A contrastive approach to the analysis of politeness formulas. *Applied Linguistics*, 8(1):75–88.
- Said Hassan Farahat. 2009. *Politeness phenomena in Palestinian Arabic and Australian English: A cross-cultural study of selected contemporary plays (PhD thesis)*. Australian Catholic University, Australia.
- Ellen Feghali. 1997. Arab cultural communication patterns. *International Journal of Intercultural Relations*, 21(3):345–378.
- A. Furnham. 1990. Language and personality. In H. Giles and W. Robinson, editors, *Handbook of Language and Social Psychology*, pages 73–95. Wiley.
- Mohammed Ghawi. 1993. Pragmatic transfer in Arabic learners of English. *El Two Talk*, 1(1):39–52.
- A. Gill and J. Oberlander. 2002. Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 363–368.
- A. Gill, A. Harrison, and J. Oberlander. 2004. Interpersonal: Individual differences and interpersonal priming. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 464–469.
- E. S. Glenn, D. Witmeyer, and K. A. Stevenson. 1977. Cultural styles of persuasion. *International Journal of Intercultural Relations*, 1(3):52–66.
- Samuel D. Gosling, Peter J. Rentfrow, and Jr. William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37:504–528.
- W. B. Gudykunst and S. Ting-Toomey. 1988. *Culture and interpersonal communication*. Sage, Newbury Park, CA.
- S. Haywood, M. Pickering, and H. Branigan. 2003. Cooperation and co-ordination in the production of noun phrases. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, pages 533–538.
- Michael Huspek. 1989. Linguistic variability and power: An analysis of you know/I think variation in working-class speech. *Journal of Pragmatics*, 13(5):661–683.
- Oliver P. John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Lawrence A. Pervin and Oliver P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138.
- Robin Lakoff. 1973. Language and woman’s place. *Language in Society*, 2(1):45–80.
- Francois Mairesse and Marilyn Walker. 2008. Trainable generation of big-five personality styles through data-driven parameter estimation. In *Proc. of 46th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30:457–500.
- Maxim Makatchev, Min Kyung Lee, and Reid Simmons. 2009. Relating initial turns of human-robot dialogues to discourse. In *Proc. of the Int. Conf. on Human-Robot Interaction (HRI)*, pages 321–322. ACM.
- Maxim Makatchev, Imran Aslam Fanaswala, Ameer Aymaan Abdulsalam, Brett Browning, Wael Mahmoud Gazzawi, Majd Sakr, and Reid Simmons. 2010. Dialogue patterns of an arabic robot receptionist. In *Proc. of the Int. Conf. on Human-Robot Interaction (HRI)*, pages 167–168. ACM.
- M. McPherson, L. Smith-Lovin, and J. M. Cook. 2000. What is a language community? *American Journal of Political Science*, 44(1):142–155.
- Clifford Nass, Y. Moon, B. Fogg, and B. Reeves. 1995. Can computer personalities be human personalities? *Journal of Human-Computer Studies*, 43:223–239.
- Gaylel Nelson, Mahmoud Al-Batal, and Erin Echols. 1996. Arabic and english compliment responses: Potential for pragmatic failure. *Applied Linguistics*, 17(4):411–432.
- R. Patai. 1983. *The Arab mind*. Charles Scribner’s Sons, New York.
- J. C. Pinheiro and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. Springer.
- Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers?: shifting demographics in mechanical turk.

- In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 2863–2872, New York, NY, USA. ACM.
- Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.
- Klaus R. Scherer. 1972. Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, 40:191–210.
- John Searle. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Jenny Thomas. 1983. Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2):91–112.
- Jenny Thomas. 1984. Cross-cultural discourse as 'unequal encounter': Towards a pragmatic analysis. *Applied Linguistics*, 5(3):226–235.

Greeting	Question-Answer	Disagreement	Apology
A: Good morning. B: What's up? Need anything?	A: Could you tell me where the library is? B: Just go to the end of the hallway, you can't miss it.	A: Could you tell me where the library is? B: Go to the second floor. A: I thought it was on the first floor. B: No, honey, there is none on the first floor.	A: Could you tell me where the library is? B: Sorry about that, I have no idea.
A: Good morning. B: Good morning. How may I help you?	A: Could you tell me where the library is? B: It's at the end of the hallway on your left.	A: Could you tell me where the library is? B: It's on the second floor. A: I thought it was on the first floor. B: No, there is no library on the first floor.	A: Could you tell me where the library is? B: Sorry, I don't know.
A: Good morning. B: Good morning, sir (madam). Would you allow me to help you with anything?	A: Could you tell me where the library is? B: Kindly follow this hallway and you will encounter the entrance on your left.	A: Could you tell me where the library is? B: Yes, you may find the library on the second floor. A: I thought it was on the first floor. B: I am afraid that is not correct, there is no library on the first floor.	A: Could you tell me where the library is? B: I have to apologize, but I don't know.

Table 2: Stimuli that correspond to negative (top row), neutral (middle row), and positive (bottom row) formality.

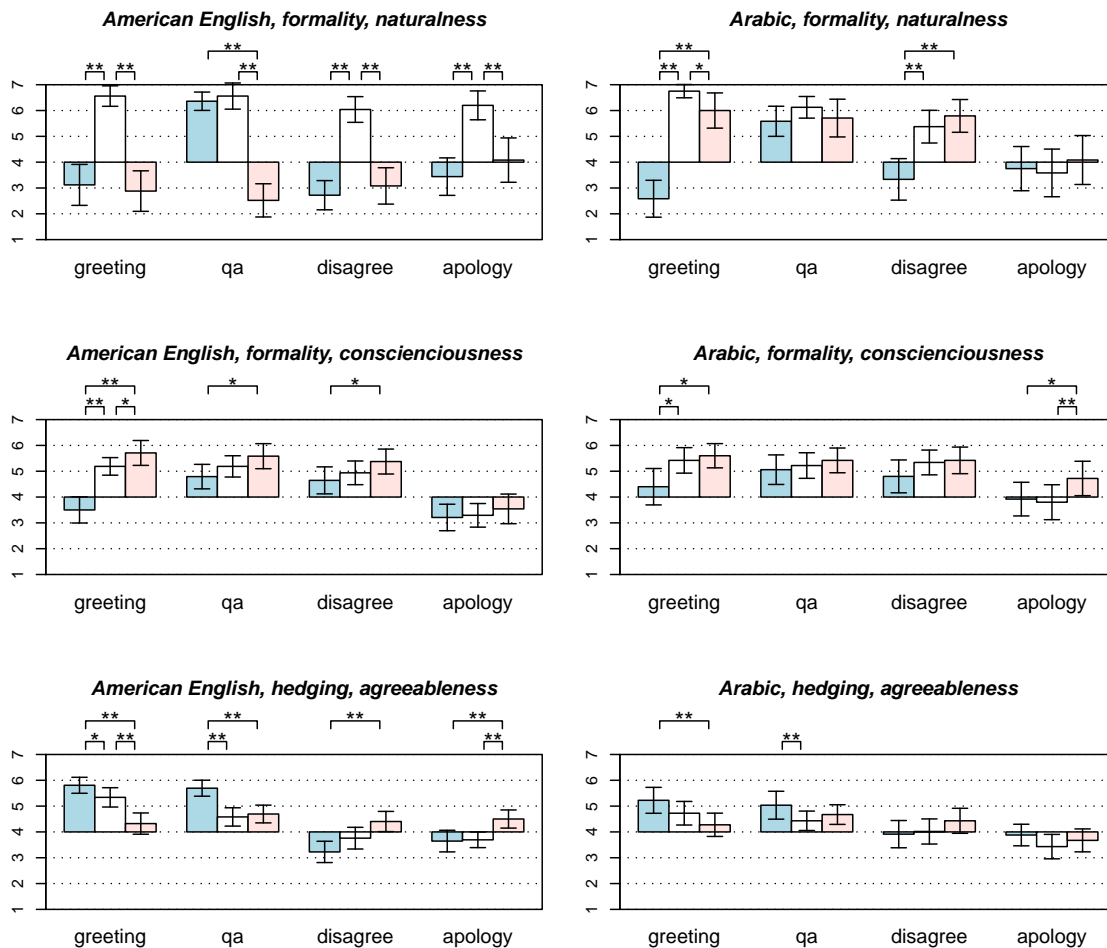


Figure 1: A subset of data comparing scores on the Big Five personality traits and naturalness as given by native speakers of American English (left half of the page) and Arabic (right half of the page). Blue, white, and pink bars correspond to negative, neutral, and positive valences of the linguistic features respectively. Dialogue acts listed along the horizontal axis are a greeting, question-answer pair, disagreement, and apology. Error bars the 95% confidence intervals, brackets above the plots correspond to p-values of paired t-tests at significance levels of 0.05 (*) and 0.01 (**)