Gender differences in the language of the Map Task dialogues

Carl Vogel
Trinity Centre for Computing
and Language Studies
Computational Linguistics Group
School of Computer Science and Statistics
Trinity College Dublin, Ireland
Email: vogel@tcd.ie

Milena Ribeiro Lopes
Trinity Centre for Computing
and Language Studies
Computational Linguistics Group
Computer Science and Statistics
Trinity College Dublin, Ireland
Email: lopesm@tcd.ie

Anna Esposito
Università della Campania "Luigi Vanvitelli"
Department of Psychology
and IIASS, Caserta, Italy
Email: iass.annaesp@tin.it

Abstract—A number of studies have revealed differences in distributions of linguistic features in the language produced by males and females and for a variety of text types (formal and informal prose, dialogue). A typical finding is a greater proportion of pronoun use among women than among men. Recently, Hans-Joerg Schmid analyzed the Human Communication Research Center (HCRC) Map Task dialogues as a context in which topic of discussion is controlled and reports variations for a number of individual items. We report here on our own exploration of the HCRC Map Task data examining categories of lexical items, and in relation to the factors that are supported by the experimental design in the construction of the Map Task: speaker gender, partner gender, prior interpersonal familiarity, task-based role, eye-contact. We find that some of the trends that have been discussed in the past depend on interactions of these factors.

I. Introduction

People appear to have conflicting intuitions about whether hypothesized differences between women and men in language use might withstand empirical scrutiny. However, empirical research has identified a number of linguistic variables that exhibit gender effects. One might imagine gender effects to be strongly influenced by social factors, and therefore it is necessary to attempt to control those. For example, women and men might have socially induced dispositions to speak or write about different topics. The text available for analysis of possible gender effects is generally that which is produced for public consumption, and is thereby subject to independent influences that merely coincide with gender distinctions. Such considerations have led to research that analyzes relevant linguistic variables while controlling for topic [1]. Gender differences that reverse in particular communication contexts have been reported [2], where contexts included conversational data and fiction. Of 37 categories for which significant gender differences emerged, 19 were stable (differing in magnitude) across conditions, while 18 showed reversed effects of gender (an *effect switch*), generally depending on the context. Pronoun use emerged as a distinguising feature, and pronoun categories were not among the reversals, thereby forming a robust gender difference. It has been suggested that use of first and second person pronouns demonstrates communicative "involvement" and that females display involvement more than males [3].

Among formal texts, fiction and non-fiction, types of pronoun use that differ significantly between male and female authors have been identified [4], women using first-person singular pronouns more than men in both fiction and non-fiction, and first-person plural pronouns more than men in non-fiction (but men using more first-person plural pronouns than women in fiction). *A priori*, first-person plural pronouns encode more inclusive involvement than first-person singular pronouns. For the entire category of first-person pronouns, women use more than men in both fiction and non-fiction. While it could be that in fiction and non-fiction separate factors influence the choice of scenarios and topics that shape pronoun preferences [1], other work [4] has demonstrated with samples selected from controlled text types that the choice to use pronouns is a stylistic preference and not forced by the domain.

That the context of the data source can influence gender discriminability is not evidence against the "gender as culture" hypothesis (see [5]); however, this depends on the granularity of context circumscription. Where reversal of a gender effect is conditioned on context, it could be that the context has culturally salient properties. Reversals between conversation and fiction are different than reversals within conversation depending on whether there is eye-contact. In communities where English is the majority language, there is not an established case that availability of eye-contact to interlocutors in dialogue makes a gender-related difference to the language used. However, the availability of eye-contact affords a non-linguistic means of establishing involvement, namely, availing of eye-contact. Thus, it is useful to explore the interaction among gender, pronoun use and availability of eye-contact. We re-examine the HCRC Map Task corpus attending to linguistic variables that have been identified as exhibiting influences that depend on whether the author (speaker) is male or female.[1]

II. Related work

A. The HCRC Map Task

The HCRC Map Task [6] data contains transcripts of 128 dialogues collected through a design that matched 64 participants (32 male, 32 female; median age 19, sd = 2.24; 61 of the

---

[1]This is actually a sex (rather than gender) difference. This paper uses "gender" in the manner of European Commission gender policy texts. See http://ec.europa.eu/justice/gender-equality/ – last verified, June 2017.

64 from Scotland; all students of the University of Glasgow) sometimes with someone known to them (participants arrived to the recordings with a friend) and sometimes with someone unfamiliar. The Map Task involves an informational asymmetry in that the two participants in a dialogue have related maps (one of 16 possible maps, with landmarks designed to elicit certain phonological and pragmatic effects (e.g. contrastive focus)), with one participant acting in the ROLE of information giver and the other as information follower: the map of the former has a path drawn upon it, and the map of the latter lacks this path. The task is for the information giver to communicate to the information follower (without either seeing the other's map) sufficiently to allow the information follower to draw the same path. The task is complicated by the fact that the maps assigned to each role differ in some of the landmarks. Half of the participants had the possibility of making eye-contact, and the other half lacked eye contact. Within the two levels of the EYE-CONTACT variable, participants twice communicated with someone familiar and twice with someone unfamiliar. This creates the FAMILIARITY variable. As each participant participated in four dialogues, EXPERIENCE is an implicit variable. Using meta-data associated with the dialogue source,[2] each turn in the dialogue may be noted with speaker GENDER, ROLE and partner gender match (GMATCH).

B. Gender-based analysis of the Map Task data

The research most directly related to our study is recent work by Hans-Jörg Schmid [1]. He asks [1, p. 329]: "Do women and men use selected words with different frequencies of occurrence if the variable TOPIC is kept constant and other variables affecting language use are also controlled?" The question is focused on relative frequency of a number of lexical types: *and*, *I*, *mmhmm*, *of*, *okay*, *the*, *you*. For each of those types, the difference in relative frequency between the four categories of Map Task dyads (female instruction giver, female instruction follower; male instruction giver, male instruction follower) is insignificant, except for *okay* and *mmhmm*. The analysis does not reveal a gender effect for either *I* or *you*, but effects of ROLE and FAMILIARITY are reported for *I* and an effect of ROLE is noted for *you*. Gender effects are noted for: *mmhmm* (females using the form more – this is interpreted in relation to conversational involvement), *and* (females using discourse-continuation *and* more, with significant interactions of the other variables), and *of* (females using the hedging form (e.g. *sort of*) more; males using the numeral (e.g. *two thirds of* and cardinal direction (e.g. *northwest of*) forms more). The paper continues to explore using regression models for the target variables individuated by relative frequency of the lexical types also taking into account other variables controlled in the HCRC Map Task data: ROLE, EYE-CONTACT, FAMILIARITY. Thus, the answer to the research question posed is [1, p. 340] "a cautious 'yes, to some extent gender-related variation continues to have an effect on language use, even when topic and virtually everything else is controlled'."

C. Unresolved questions

The data preparation method of the work reported in §II-B [1] divided each of the 128 transcript files in two, one file for each participant in the transcribed conversation. The method

of analyses is based on relative frequencies of the seven items mentioned. An alternative is to analyze the conversations by turns within them, and treating the data as counts and proportions (counts of items of interest in relation to counts of complementary items). Treating each turn as a point of measurement yields 27084 cases. The number of turns, as a function of gender and dialogue/task role, is shown in Table I; the same breakdown for total tokens is shown in Table II.[3] Table III indicates the total number of tokens per turn, by gender, and the central tendency does not differ between genders. Table IV shows tokens per turn, by role.

TABLE I.   TOTAL TURNS BY GENDER AND ROLE

| Gender | Dialogue Role | | Total |
	Information Follower	Information Giver	
Female	5844	7732	13576
Male	6028	7480	13508
Total	11872	15212	27084

TABLE II.   TOTAL TOKENS BY GENDER AND ROLE

| Gender | Dialogue Role | | Total |
	Information Follower	Information Giver	
Female	21550	50876	72426
Male	26385	54969	81354
Total	47935	105845	153780

TABLE III.   SUMMARY STATISTICS OF TOTAL TOKENS SPOKEN PER TURN, ACCORDING BY GENDER

Speaker Gender	median	mean	s.d.	max	min
Female	3	5.3	6.1	59	1
Male	4	6.0	6.4	106	1

TABLE IV.   SUMMARY STATISTICS OF TOTAL TOKENS SPOKEN PER TURN, ACCORDING BY ROLE

Speaker Role	median	mean	s.d.	max	min
Information Follower	2	4.0	4.5	49	1
Information Giver	5	7.0	7.3	106	1

For the 128 dialogues, the maximum number of turns is 685 (Minimum = 42; median = 183 (Mean = 211.6, s.d. = 107.7)). A χ^2 test of independence between total number of turns and gender is not significant, but between number of tokens uttered and gender, is significant ($\chi^2 = 518.33, df = 1, p < 0.001$), with males producing more tokens than females. The interaction between total number of turns and role is also significant ($\chi^2 = 411.89, df = 1, p < 0.001$), with information givers taking more turns than information followers. The interaction between total tokens produced and role is significant ($\chi^2 = 21808, df = 1, p < 0.001$), information givers uttering more tokens than information followers. A log-linear model of the interaction between gender and role on total tokens obtains Pearson residuals with absolute values greater than 4 ($p < 0.01$) for each cell:[4] females used fewer tokens than expected in the role of information follower and more tokens than expected in the role of information giver; males used

[3]To profile the volume of data, the token count is not of just the seven types from previous analysis [1], nor just the specific items of interest for our study, as indicated in Table V, but total number of tokens spoken in the dialogues.

[4]Residuals with absolute value between 2 and 4 are significant with 95% confidence, and greater than 4, with 99% confidence.

fewer tokens than expected in the role of information giver and more tokens than expected in the role of information follower. The Pearson residuals are not significant in the interactions between gender and role on total turns.

These descriptive values indicate patterns in the interactions that imply significant gender differences in total turns and total number of tokens uttered, and, further, that other variables controlled within the HCRC Map Task experiment (e.g. task-based role) interact with gender. This paper seeks to explore questions about gender differences in the HCRC Map Task data connected to the use of pronouns and in relation to variables controlled within the original experiment that may be expected to interact with pronoun use. As described above, use of pronouns has been interpreted as an indication of perceived mutual involvement in conversation (since pronoun use presupposes shared reference resolution), and it has been suggested that females are more likely to display conversational involvement. However, this may interact with the gender of the partner and the familiarity of the partner. It may also interact with experience with the task itself. It was noted above that [4] found females to use first-person singular pronouns more than males in both fiction and non-fiction, but that first-person plural pronouns depended on genre: females used them more than men in non-fiction, but men using more first-person pronouns than women in fiction.[5] If the HCRC Map Task data follows this trend, then the data should pattern with non-fiction, since even though the maps depict fictional locations, the discussion is of physical maps and the markings upon them.

III. METHOD AND QUESTIONS OF FOCUS

The 128 plain text files of the HCRC Map Task were processed by turn (each is annotated as of the information giver or information follower), and the number of instances of a range of lexical types (here we focus on pronouns, see Table V) per line were counted, along with the total tokens per line.[6] Each of these counts was aligned with the meta-data associated with the dialogue. This includes participant data (gender, age, etc) and dialogue data – the role of each participant, whether eye-contact was available, the map used, the ordinal position of the task for the participants in their experience with the task, the ultimate task success score (path deviation), etc.. Given that there are differences in total tokens produced in relation to both gender and task role, analyses are conducted relative to these factors. To demonstrate further effects we use association plots of log-linear models, using the significance of Pearson standardized residuals in order to assess interactions that are significant [7], [8]. Given that the data is ultimately represented as count proportions, logistic regression would also be appropriate.

We seek answers to the following questions, with respect to the controlled conditions of the HCRC Map Task experiment:

---

[5]For the entire category of first-person pronouns, women used more than men across genres.

[6]In Table V, one may note omissions: *theirselves* and *theirself* – these are not attested in the corpus. An item classified here as interrogative (*however*) is a homonymn of a discourse connective; however, the only attested use in the corpus of *however* is as relative pronoun, which is frequent use of interrogatives: "just go round go round the overgrown gully however you want then just end up above safari truck" ("however you want" is a manner-adverbial relative clause). The corpus is provided with entirely lower-case letters; therefore, the wordlist is presented in the same way (*i* vs *I*).

Type	Instances
DEMONSTRATIVE	*this, that, these, those*
INTERROGATIVE	*who, whose, which, whichever, whom, whomever, how, however, when, whenever, where, wherever, what, whatever, why*
FIRST PERSON SINGULAR	*i, me, myself, my, mine*
FIRST PERSON PLURAL	*we, us, our, ours, ourselves*
Second person	*you, your, yours, yourself, yourselves*
THIRD PERSON	*they, them, their, theirs, it, he, she, him, her, his, hers, herself, himself, themselves, itself, its,* DEMONSTRATIVES, INTERROGATIVES

1) Does dialogue role interact with gender and the use of *I* and *you* (cf. [1], as discussed in §II-B)?

2) Does dialogue role interact with gender and the use of pronouns as a broad category (personal, demonstrative, reflexive, interrogative) (cf. [3], per §I)?

3) Does dialogue role interact with gender and the use of first person singular and plural pronouns (cf. [4], as discussed in §I)?

4) Are there mitigating effects of the availability of eye-contact, familiarity of partners in a dyad, partner gender or task-experience?

IV. RESULTS

A. I and you

The residuals are not significant for the interaction between the count of *i* (vs. the total count of other pronouns per turn nor vs. the count of other tokens per turn) and gender. However, the interaction with role is significant (in relation to both other pronouns and tokens per turn): information followers use *i* more than would be expected and information givers use *i* less than would be expected without interaction ($p < 0.01$). The interaction is depicted in the association plot of Fig. 1. Association plots indicate the magnitude of the difference (residuals) between observed values and values that would be expected if there were no interaction among cross-tabulated variables. Areas above the dotted line indicate observations that exceed expectations, and areas below the line indicate smaller than expected observations. The darkness of shading corresponds to significance levels.[7] In the case of Fig. 1, a proportion is constructed (IvTP) – counts of *i* in relation to counts of all other pronouns (Noti), and this proportion is shown in relation to both role and gender.

There is a significant interaction between gender and the use of *you* (relative to both all other pronouns and all other tokens, $p < 0.05$; females used *you* more and males less than expected if there were no interaction between gender and use of *you* – see Table. VI). This contrasts with the result reported by [1], with aggregation over entire dialogues, rather than by turn, as analyzed here. There is also a significant effect of role (followers using *you* less and information givers using *you* more than expected if there were no interaction, $p < 0.01$ – see Table VII). Fig. 2 illustrates a gender effect reversal for information givers depending on eye-contact, with females using *you* more than expected without eye-contact (and males, less) and males using *you* more than expected with eye-contact

---

[7]Tables VI and Tables VII present similar information as contingency tables to which χ^2 tests are applied: *italic* corresponds to significantly smaller than expected values and **bold**, to significantly larger than expected values.
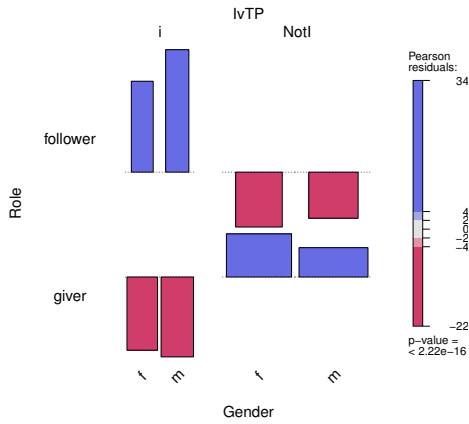
Fig. 1. Interaction among use of *I* vs. total of other pronouns (IvTP), task role, and gender (NotI corresponds to pronouns that are not *I*)

(and females, less), in relation to the expectations that would be in place without an interaction ($p < 0.01$). No such reversal obtains in relation to partner gender, familiarity or task-based experience.

TABLE VI. TOTAL COUNT OF *you* VS. OTHER TOKENS (YOUvTT) IN RELATION TO GENDER (PERSON RESIDUALS IN PARENTHESES)

Gender	YouvTT				Total
	you		not *you*		
Female	**3243**	**(2.336)**	69183	(-0.495)	72426
Male	*3366*	*(-2.204)*	77988	(0.467)	81354
Total	6609		147171		153780

TABLE VII. TOTAL COUNT OF *you* VS. OTHER TOKENS (YOUvTT) IN RELATION TO ROLE (PERSON RESIDUALS IN PARENTHESES)

Role	YouvTT				Total
	you		not *you*		
Follower	*618*	*(-31.773)*	**47317**	**(6.733)**	47935
Giver	**5991**	**(21.382)**	99854	(-4.531)	105845
Total	6609		147171		153780
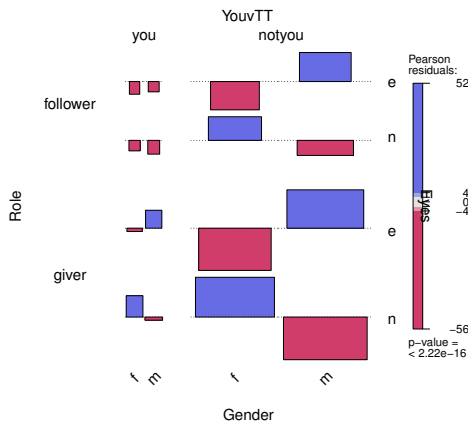
Fig. 2. Interaction among use of *You* vs. total other tokens (YouvTT), gender, role and availability of eye contact

B. Pronouns vs NonPronouns

There is not a significant difference in the total use of pronouns (in relation to other tokens) by males and females, but there is a significant effect between pronoun use and dialogue role ($\chi^2 = 76.89, df = 1, p < 0.001$), with followers using pronouns more and information givers using pronouns less than would be expected if there were no interaction ($p < 0.01$; see Fig. 3) Fig. 4 shows that there is an interaction ($p < 0.01$) among pronoun use, dialogue role, gender and eye-contact, with gender effect reversal in relation to availability of eye-contact: females use pronouns more than expected (if there is no interaction), and males less, when there is no eye contact, and males use pronouns more than expected (and females less) when there is eye contact. Unlike the situation for *you* as an individual word where the contrast holds only within the information giver role (see §IV-A), for the entire category of pronouns, the reversal holds for both roles. The effect switch does not obtain in relation to matching of partner gender as it does with eye-contact, but it does obtain in relation to experience within the information follower role (both men and women use more total pronouns than would be expected if there were no interaction ($p < 0.01$) in their first of four experiences with the task; men use more pronouns than expected ($p < 0.05$) in the second game; women use fewer ($p < 0.05$) and men use more ($p < 0.01$) pronouns than expected in the third game; women use fewer ($p < 0.05$) in their fourth game). Familiarity also involves a switch (see Fig. 5): as information followers, females and males whose partner is familiar use more pronouns than expected, just as for information followers without taking into account gender or familiarity (Fig. 3), but where the partner is not familiar, females use fewer total pronouns than expected as information follower ($p < 0.01$), while males use more ($p < 0.05$).
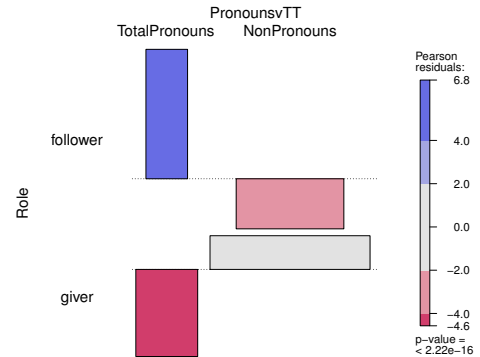

Fig. 3. Interaction among use of pronouns vs. total other tokens (PronounsvTT) and role

C. First person pronouns

1) First person pronouns: There is an interaction between gender and first person pronouns (relative to the remaining tokens) ($\chi^2 = 15.09, df = 1, p < 0.001$), with females using fewer first person pronouns and males using more than would be expected if there were no interaction ($p < 0.05$). The same effect is visible (see Fig. 6) if first person pronouns
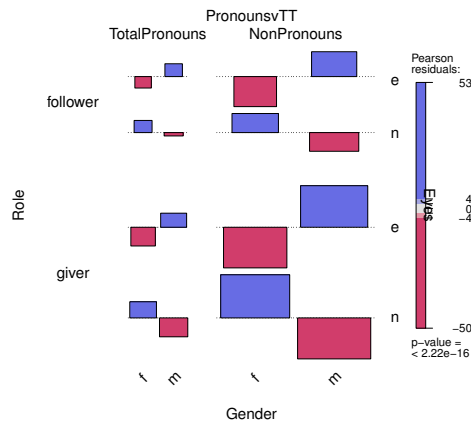
Fig. 4. Interactions: pronoun vs. total other tokens (PronounsvTT), role, gender, eye contact
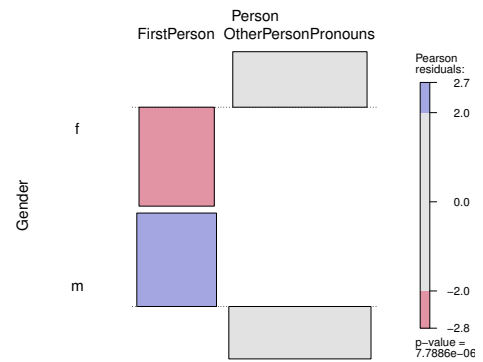

Fig. 5. Interactions: pronoun vs. total other tokens (PronounsvTT), role, gender, familiarity

are considered with respect to all other pronouns ($\chi^2 = 19.84, df = 1, p < 0.001$). In what follows, we address only the comparison with other pronouns. There is an effect of role, with information followers using more first person pronouns and information givers using fewer than would be expected if there were no interaction ($p < 0.01$; see Fig. 7).

Interactions among gender, role and other dimensions are interesting. Fig. 8 shows a reversal of the gender effects in interaction with the availability of eye contact: females use fewer first person pronouns than expected with the availability of eye contact and more than expected without the availability of eye contact ($p < 0.01$), while males use more first person pronouns than expected with the availability of eye contact and fewer than expected without the availability of eye contact. In contrast, the interaction between first person pronoun use and gender does not reverse in interaction with partner gender match (rather, the main effect is preserved only in the case of gender-mixed dyads). Fig. 9 shows an alternative interaction among first person pronoun use, gender and familiarity: females use fewer first person pronouns than expected with unfamiliar partners (than if there were no interaction, $p < 0.05$); males use more first person pronouns than expected with familiar partners (than if there were no

Fig. 6. Interactions: first person pronoun (vs. other pronouns) use and gender

Fig. 7. Interactions: first person pronoun (vs. other pronouns) use and role

interaction, $p < 0.05$). Finally, it is also evident that there is an effect of task experience (Fig. 10): evidently, the main gender effect (Fig. 6) is dominated by the fewer than expected first person pronouns used by females with increased experience, and the more than expected use of first person pronouns in the first experience of the task for males.
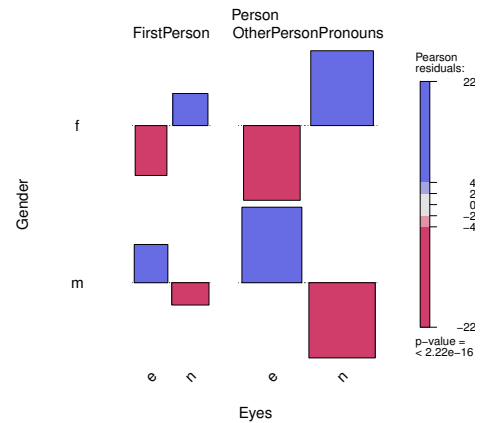

Fig. 8. Interaction among use of first person pronouns (vs. other pronouns), gender and availability of eye contact
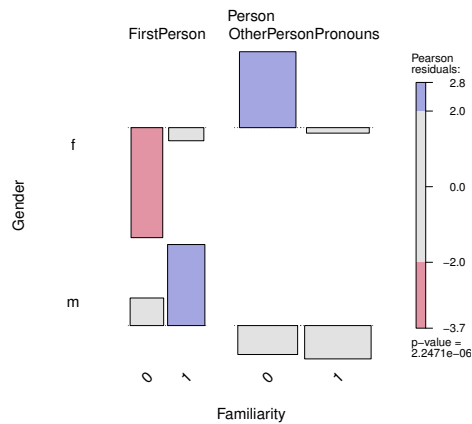
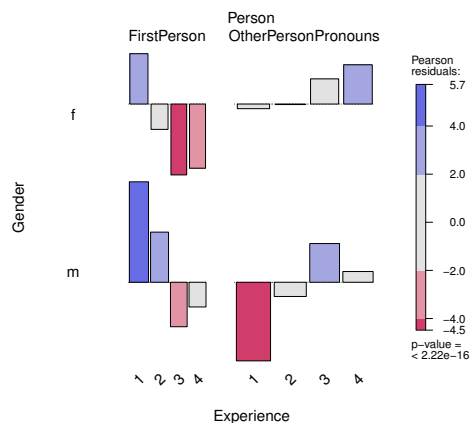Fig. 9. Interactions among use of first person pronouns (vs. other pronouns), gender and familiarity

Fig. 10. Interaction among use of first person pronouns (vs. other pronouns), gender and task experience

2) First person singular: The main effects of first person singular pronouns in interaction with gender and with role replicate those of the larger category of first person pronouns.

3) First person plural: The main effects of first person plural pronouns in interaction with gender replicate those of the larger category of first person pronouns; however, the interaction between role and use of first person plural pronouns does not achieve significance.

V. DISCUSSION AND CONCLUSIONS

Our analysis of the HCRC Map Task data takes measurements at finer level of granularity than that reported in an earlier study of gender effects within it [1]. We replicate the findings of the earlier study for *I*, but identify significant effects of gender for *you*. Moreover, interactions among role, gender and availability of eye-contact exist. While there is not a gender effect in the use of the broad category of pronouns (along the lines of women using more pronouns than men), role matters, and gender interacts with dialogue role, familiarity and availability of eye contact. First person singular and plural pronouns pattern largely together, but they differ in interactions with role. The lack of eye-contact condition

is shared with study of formal texts reported above [4],[8] and without eye-contact in dialogue females used more pronouns, just as in prose. These results may be understood in relation to signaling involvement: linguistic effort to signal involvement is superfluous where eye-contact is available.

This work supports advances in modeling human-machine interaction and the quest to develop algorithms, and computational paradigms to implement culture-specific, trustful, satisfactory and emotionally coloured human-machine interfaces [9]. Machines must be endowed of knowledge and practical abilities to render the world interpretable while interacting with users, developing complex forms of human-machine entanglements [10]. Language analyses for detecting social roles and contexts will yield new cognitive theories to identify and link to context and social signals, feeding this information into mathematical models of embodiment.

REFERENCES

[1] H.-J. Schmid, "Does gender-related variation still have an effect, even when topic and (almost) everything else is controlled?" in *Change of Pardigms–New Paradoxes: Recontextualizing Language and Linguistics*, ser. Applications of Cognitive Linguistics, J. Daems, E. Zenner, K. Heylen, D. Speelman, and H. Cuyckens, Eds. Berlin: De Gruyter Mouton, 2015, vol. 31, pp. 327–346.

[2] M. L. Newman, C. J. Groom, L. D. Handelman, and J. W. Pennebaker, "Gender differences in language use: An analysis of 14,000 text samples," *Discourse Processes*, vol. 45, no. 3, pp. 211–236, 2008.

[3] D. Biber, S. Conrad, and R. Reppen, *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, 1998.

[4] S. Argamon, M. Koppel, J. Fine, and A. R. Shimoni, "Gender, genre, and writing style in formal written texts," *Text*, vol. 23, no. 3, pp. 321–346, 2003.

[5] A. Mulac, J. J. Bradac, and P. Gibbons, "Empirical support for the gender-as-culture hypotheses: An intercultural analysis of male/female language differences," *Human Communication Research*, vol. 27, no. 1, pp. 121–152, 2001.

[6] A. H. Anderson, M. Bader, E. G. Bard, E. H. Boyle, G. M. Doherty, S. C. Garrod, S. D. Isard, J. C. Kowtko, J. M. McAllister, J. Miller, C. F. Sotillo, H. S. Thompson, and R. Weinert, "The hcrc map task corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1992.

[7] M. Friendly, "Visualizing categorical data," in *Cognition and Survey Research*, M. Sirken, D. Hermann, S. Schechter, N. Schwartz, J. M. Tanur, & R. Tourangeau, Eds. John Wiley & Sons, 1999, pp. 319–348.

[8] D. Meyer, A. Zeileis, and K. Hornik, "The strucplot framework: Visualizing multi-way contingency tables with VCD," *Journal of Statistical Software*, vol. 17, no. 3, pp. 1–48, 2006.

[9] A. Esposito, A. M. Esposito, and C. Vogel, "Needs and challenges in human computer interaction for processing social emotional information," *Pattern Recognition Letters*, vol. 66, pp. 41–51, 2015.

[10] P. Baranyi and A. Csapo, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

---

[8] None of FAMILIARITY, GMATCH, EXPERIENCE map onto the fiction/non-fiction distinction.