Gender differences in verbal behaviour in a call routing speech application

Håkan Jonsson* & Robert Eklund*†‡
* Voice Provider Sweden AB
† Linköping University, Department of Computer Science
‡ Karolinska Institute, Department of Cognitive Neuroscience

Abstract

This paper reports results on verbal behavior in a live natural language call routing speech application. Differences between male and female callers in terms of verbosity are investigated, and put in relation to three variations of the system prompts. Findings show that in this particular application female callers are more verbose than male callers for open style prompts, while there is no difference for a directed style prompt.

Introduction

Almost all organisations and companies that handle large volumes of incoming telephony contacts from customers and end users have a need for some kind of call routing, that is, some means of assuring that calls reach their proper destinations promptly and with a high degree of service. As businesses and organisations grow in complexity, call routing has become a common application of speech technology. In many cases, such applications ask so-called open-ended questions, using natural language understanding technologies to allow callers to speak freely to describe their reason-for-calls.

This paper reports on findings in behavioural differences between genders in one such system, handling incoming calls to a major Swedish retail company.

Background

The study of open-ended call routing applications, oftentimes referred to as How May I Help You-systems, or HMIHYs, started with Gorin et al. (1997). They investigated data-driven methods for the development of such systems, and to date similar methods remain dominant, though in different flavours and with different means of representing meaning (cf. Kituno et al., 2003; Huang & Cox, 2006; Boye & Wirén, 2007; Lee et al., 2000, for a general discussion).

Behavioral aspects in call routing have been investigated since the late 1990s. For example, McInnes et al. (1999) found that using more 'open' style questions when prompting the caller for input elicited longer responses than when using a more direct, or 'closed' prompt style. Sheeder & Balogh (2003) investigated the impact on responses when the system presented examples of typical user responses before prompting the caller to speak, showing that examples increased routing precision in the application. Williams & Witt (2004) explored prompting strategies in relation to the domain of the application. They found that in a domain where callers had clear expectations on the structure of the task, open style prompting was more successful than in a domain where caller expectations were more vague. Eklund & Wirén (2010) investigated prompt style and its effects on the presence of filled pauses in user responses. They found that prompt style greatly influenced callers' verbosity in that open style prompts elicited longer answers.

None of these studies focused on caller characteristics such as gender, despite the fact that gender has most likely played a central role in human history. Detailed caller characteristics may often be inaccessible for analysis, but gender stands out in that it can usually be determined from just listening to recordings, with at least some degree of certainty.

The exact role, function and status the genders have been "given" has varied a lot throughout history, and recent Western political debate has highlighted gender aspects.

Method

The data we present in this study were collected from live callers using a prototype call routing application. The prototype application was designed solely for the purpose of data collection. It did not employ any real routing

logic, but posed as a live application in order to elicit genuine responses from callers.

The application was deployed and integrated with the main call center of a major Swedish retail company. Hence, all data are collected from real-world customers, calling in with real reason-for-calls. The original goal of this setup was to analyze and assess the viability of a full-functionality call routing application. The data utilized here is a subset of the original data set.

The prototype application, having no internal logic, presented each caller with two prompts; first an initial *main prompt*, and then one *follow-up prompt*.

For the main prompt, there were three alternative prompt wordings, or prompt styles, as follows:

Open question prompt:
  "How may we help you?"
  ("Vad kan vi hjälpa dig med")

Basic question prompt:
  "What is the reason for your call?"
  ("Vad gäller ditt ärende?")

Keyword prompt:
  "Please state, with a word or two, the reason for your call!"
  ("Säg med ett eller två ord vad ditt ärende gäller")

The intention was to achieve variation in prompt style in terms of open-endedness. In any given call, the prompt was selected pseudo-randomly, even though the distribution is not equal – the basic question prompt was played in 50% of the calls, the other two in 25% each.

Follow-up prompts were also varied in a corresponding fashion, independently of the main prompts. However, effects of this variation are not detailed in this study.

The recorded utterances were transcribed and annotated according to the following classifications:

Gender – male or female. This was judged according to the annotators' perception only.

Number of words. The number of full lexical items uttered, not non-verbal sounds, etc.

Informativeness. Informative or non-informative. Whether the utterance contained any information pertaining to the subject matter of the call.

Finite verb. If there were was one (or more) finite verb uttered in the utterance.

In total, 856 calls, from 363 and 493 male and female callers, respectively, were transcribed and annotated.

Results

Table 1 shows the average number of words spoken in utterance 1 and utterance 2 for male and female callers, respectively. Note that no differentiation is made for prompt style.

Table 1. Average number of spoken words per utterance

Gender	Utt 1	Utt 2
Male	1.66	1.72
Female	2.52	2.26

As can be seen, females utter more words in both the first utterance ($p = 0.000074$, *t*-test, two-tailed) and in the second utterance ($p = 0.0066$, *t*-test, two-tailed).

Investigating whether this difference holds for all prompt styles, *Table 2* shows the average number of words for utterance 1, as a function of prompt style.

Table 2. Average number of spoken words in the first utterance, for each prompt style

Gender	Open	Basic	Keyw.
Male	2.03	1.65	1.29
Female	3.51	2.71	1.36

As is seen here, female callers utter more words than male callers in response to both the open prompt ($p = 0.0029$, *t*-test, two-tailed) and the direct prompt ($p = 0.00096$, *t*-test, two-tailed). However, for the keyword prompt, the difference is not significant ($p = 0.64$, *t*-test, two-tailed). This suggests that prompt style influences male and female callers' verbal behaviour to different degrees.

To investigate the nature of these differences further, the presence of finite verbs in responses was investigated (see Eklund & Wirén, 2010). The assumption was that finite verbs would be an indicator of more conversational responses. The results are shown in *Table 3*.

Table 3. Proportion of utterances containing a finite verb, per utterance

Gender	Utt 1	Utt 2
Male	12.9%	11.8%
Female	20.5%	15.4%

As is seen, females are more likely to include a finite verb form than are males ($p = 0.0053$, Z-test, two-tailed). There is a similar tendency for the second utterance, but the difference is not significant ($p = 0.16$, Z-test, two-tailed).

In *Table 4* it can be seen that female callers produce more utterances that contain at least one finite verb than male callers do for the open prompt ($p = 0.016$, Z-test, two-tailed) as well as for the basic prompt ($p = 0.023$, Z-test, two-tailed). For the keyword prompt, there is no significant difference.

Table 4. Proportion of utterances containing a finite verb, per utterance

Gender	Open	Basic	Keyw.
Male	14.1%	13.7%	9.0%
Female	28.8%	23.1%	9.7%

Still, as seen in *Table 5* below, there are no significant differences in terms of informative utterances between males or females, for any of the three prompt styles.

Table 5. Proportion of utterances rated as informative

Gender	Open	Basic	Keyw.
Male	87.7%	87.9%	87.2%
Female	88.2%	87.3%	84.2%

Discussion and future work

In short, the main finding here is that in this particular application, female and male callers respond differently to prompts that are more conversational in style. The female callers' responses are more verbose, and contain more finite verbs. These differences occur when the system uses a prompt style that to a higher degree encourages such behaviour, whereas for the keyword prompt, there are no significant differences between male and female callers. It would seem that in this particular case, female callers are more influenced by, or receptive to, the "speech style" of the system.

Looking for an explanation for the observed differences, the most obvious one would perhaps be that women just "speak more than men", which is also a claim that is often encountered. The origin of this claim is most likely Louann Brizendine's (2006) bestseller *The Female Brain* where it was claimed that women, on average, use 20,000 words per day, whereas men use only 7,000 words, a claim that received a lot of attention. However, closer scrutiny revealed that these figures had no scientific ground, and several subsequent, quantitative, studies either found no differences between men and women as to verbosity (Mehl et al., 2007; Cameron, 2007; Liberman, 2006), or even found that men are in fact more verbose than women (Leaper & Ayres, 2007).

However, several studies *do* point to gender differences in both cognitive abilities in general (Halpern & Tan, 2001; Mann et al., 1990) and in linguistic style/behaviour (e.g. Leaper & Ayres, 2007; Haas, 1979; Crosby & Nyquist, 1977).

While the reported general trend is that men, as a group, have superior spatial ability (Kimura, 1996), women, as a group, tend to exhibit superior verbal fluency (Halpern & Tan, 2001; Mann et al., 2001). One possible explanation for this that has been offered is that females have a thicker corpus callosum than men (De Lacoste-Utamsing & Holloway, 1982) which could perhaps also explain the often reported claim that women have less lateralized hemispheres for several cognitive functions, including language processing (Shaywitz et al., 1995).

However, although several studies point to neurological gender differences (which would also explain gender differences concerning e.g. dyslexia etc.), there are also several studies that have found no significant differences between the sexes (Brouwer, Gerritsen & De Haan, 2007; Sommer et al., 2004; Hyde, 1981). For reviews, see Cahill (2006) and Frost et al. (1999).

Another important factor that needs to be mentioned is the voice (gender) of the *system*, which in this case was male. One can speculate that females perceive this as more authoritative, leading them to speak more in the style of the prompts they respond to. It would be a natural extension of this study to also control for the variation of the gender of the system (see Nass & Brave, 2005).

These factors aside, considering other possible explanations for these behavioural differences, one possibility could be to consider how callers perceive the system. It may be argued that female callers, to a higher degree than male callers, view the system as a *conversational partner* with capabilities similar to a human. Male callers, on the other hand, would then perceive the system more as just another user *interface*. This is consistent with Edlund et al. (2008), who propose that users' conceptions of spoken dialog systems, and their interactional

style, can be explained in terms of what type of metaphor they use to conceive the system.

Finally, we can conclude that we have observed significant differences between the genders in the interaction with this particular human–machine dialog system. Whether or not this result is a fluke, or has an underlying basis in how the genders interact with systems needs to be investigated in future research.

However, irrespective of the underlying reasons for potential language use differences between the genders, the observed difference could (perhaps) conceivably constitute a parameter to consider when designing HMIHY systems with a clear gender bias in user profiles.

Acknowledgements

References

Brizendine, Louann (2006). *The Female Brain*. New York: Morgan Road.

Boye, Johan & Mats Wirén (2007). Multi-slot semantics for natural-language call routing systems. In*: Bridging the Gap: Academic and Industrial Research in Dialog Technologies Workshop Proceedings*, 26 April 2007, Rochester, New York, 68–75.

Brouwer, Dédé, Marinel Gerritsen & Dorian De Haan (2007). Speech differences between women and men: on the wrong track? *Language in Society* 8:33–50.

Cahill, Larry (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience* 7:477–484.

Cameron, Deborah (2007). Applied linguistics and the perils of popularity. *International Journal of Applied Linguistics* 17(3):392–395.

Crosby, Faye & Linda Nyquist (1977). The female register: an empirical study of Lakoff's hypotheses. *Language in Society* 6:313–322.

De Lacoste-Utamsing, Christine & Ralph Holloway (1982). Sexual Dimorphism in the Human Corpus Callosum. *Science* 216:1431–1432.

Edlund, Jens, Joakim Gustafson, Mattias Heldner & Anna Hjalmarsson. (2008). Towards human-like spoken dialogue systems. *Speech Communication* 50(8–9):630–645.

Eklund, Robert & Mats Wirén (2010). Effects of open and directed prompts on filled pauses and utterance production. In: *Proceedings of Fonetik 2010*, Lund University, 2–4 June 2010, 23–28.

Frost, Julie A., Jeffrey R. Binder, Jane A. Springer, Thomas A. Hammeke, Patrick S. F. Bellgowan, Stephen M. Rao & Robert W. Cox (1999). Language processing is strongly left lateralized on both sexes. *Brain* 122:199–208.

Gorin, A. L., G. Riccardi & J. H. Wright (1997). How may I help you? *Speech Communication*, 23:113–127.

Haas, Adelaide (1979). Male and female spoken language differences: stereotypes and evidence. *Psychological Bulletin* 86:616–626.

Halpern, Diane F. & Uner Tan (2001). Stereotypes and Steroids: Using a Psychobiosocial Model to Understand Cognitive Sex Differences. *Brain and Cognition* 45:392–414.

Huang, Qiang & Stephen Cox (2006). Task-independent call-routing. *Speech Communication* 48:374–389.

Hyde, Janet Shibley (1981). How Large Are Cognitive Gender Differences? *American Psychologist* 36(8):892–901.

Kimura, Doreen (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current Opinion in Neurobiology* 6(2):259–263.

Leaper, Campbell & Melanie M. Ayres (2007). A Meta-Analytic Review of Gender Variations in Adults' Language Use: Talkativeness, Affiliative Speech, and Assertive Speech. *Personality and Social Psychology Reviews* 11:328–363.

Lee, Chin-Hui, Bob Carpenter, Wu Chou, Jennifer Chu-Carroll, Wolfgang Reichl, Antoine Saad & Qiru Zhou (2000). On natural language call routing. *Speech Communication* 31:309–320.

Liberman, Mark (2006). Sex on the brain. Women use 20,000 words a day, men only 7,000 – or so says a new bestseller. Fact-checking "The Female Brain". *Boston Globe*, 24 September 2006, D1.

Mann, Virginia A., Sumiko Sasanuma, Naoko Sakuma & Shinobu Masaki (1990). Sex differences in cognitive abilities. A cross-cultural perspective. *Neuropsychologia* 28(10):1063–1077.

McInnes, F. R., I. A. Nairn, D. J. Attwater & M. A. Jack (1999). Effects of prompt style on user responses to an automated banking service using word-spotting. *BT Technology Journal* 17:160–171.

Mehl, Matthias R., Simine Vazire, Nairán Ramírez-Esparza, Richard B. Slatcher & James W. Pennebaker (2007). Are Women Really More Talkative Than Men? *Science* 317:82.

Nass, C & S Brave (2005). *Wired for speech*. Cambridge. MA: The MIT Press.

Shaywitz, Bennet A., Sally E. Shaywitz, Kenneth R. Pugh, R. Todd Constable, Pawel Skudiarski, Robert K. Fulbright, Richard A. Bronen, Jack M. Fletcher, Donald P. Shankweiler, Leonard Katz & John C. Gore (1995). Sex differences in the functional organization of the brain for language. *Nature* 373:607–609.

Sheeder, Tony & Jennifer Balogh (2003). Say it Like You Mean it: Priming for Structure in Caller Responses to a Spoken Dialog System. *International Journal of Speech Technology* 6:103–111.

Sommer, Iris E. C., André Aleman, Anke Bouma & René S. Kahn (2004). Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies. *Brain* 127:1845–1852.

Speech, Music and Hearing

Quarterly Progress and Status Report

TMH-QPSR, Volume 51, 2011

Proceedings from

Fonetik 2011

June 8 – June 10, 2011