

Gender in Everyday Speech and Language: A Corpus-based Study

Diana Binnenpoorte, Christophe Van Bael, Els den Os, Lou Boves

Centre for Language and Speech Technology (CLST)
Radboud University Nijmegen, Nijmegen, The Netherlands
{d.binnenpoorte,c.v.bael,e.denos,l.boves}@let.ru.nl

Abstract

This paper presents an exploratory study on the relations between gender and everyday parlance. A “data-mining” approach is used to explore gender-specific characteristics in a large number of spontaneous telephone and face-to-face conversations. Our study focuses on speech rate (speaking rate and articulation rate), disfluencies (filled pauses and repetitions), pronunciation variation (phoneme substitutions, deletions and insertions), and preferences for particular parts of speech. Our study reveals interesting similarities and differences in everyday male and female speech, and proves that data-mining on large spoken language corpora is a promising approach for obtaining information on spontaneous speech phenomena and for generating new hypotheses for research.

1. Introduction

The increasing availability of large annotated spoken language corpora for various languages [1, 2, 3, among others] enables extensive studies of real-life spoken language phenomena. The recently compiled Spoken Dutch Corpus, also known as CGN (Corpus Gesproken Nederlands) [2], is one such corpus. Its large amount of speech data and its multi-level data annotations make the Spoken Dutch Corpus very interesting for *data-mining*, a methodological approach in which the data rather than well-defined hypotheses form the starting point for research. As such, exploratory studies can be conducted on a variety of (socio-)linguistic aspects of speech and language use. In this paper, data from the Spoken Dutch Corpus were used to study gender-related phenomena in conversational speech, a potentially important research issue for a wide range of applications.

This study aims to investigate the possible effect of a speaker’s gender on speech rate, the occurrence of disfluencies and pronunciation differences, and the potential preferences for the use of words with particular parts of speech (POS), since these variables are relatively easy to isolate and to study by means of a data-mining approach. These phenomena have been studied in experimental and corpus-based research, e.g. using the American English Switchboard corpus [4, 5, 6]. A comparison of our results with results obtained through research on American English enabled us to assess the corpus-dependency of our results. At the same time, it enabled us to assess data-mining as an explorative procedure to reveal information on speech processes and to trigger research questions for more thorough investigation.

2. Material and method

The Spoken Dutch Corpus [2] is a collection of contemporary speech as spoken in the Netherlands and in Flanders. The corpus comprises 9 million words that were uttered in different communicative settings. All recordings were orthographically transcribed, lemmatized and enriched with POS information. A 900K subset of the corpus, the so-called *core corpus*, comes with more detailed annotations, viz. a manually verified automatic transcription and a hand-checked word alignment. All transcriptions and annotations used in our study were extracted from this *core corpus*.

The data of the Spoken Dutch Corpus are distributed over 15 categories, each comprising speech recorded in a specific situational setting. In our study, we investigated the transcriptions of spontaneous speech recorded in face-to-face conversations (FTF) and telephone dialogues (TD) in the Netherlands. We successively investigated the speech rate, the occurrence of fillers and repetitions, the pronunciation variation and the use of particular parts of speech in male and female speech. Direct comparisons between the male and female speech are presented for each variable. Whenever appropriate, details on the impact of the interlocutor’s gender on speech are presented as well. Given the exploratory nature of this study, our research was limited to the processing of the data annotations, so no signal processing was involved.

Table 1 presents an overview of the data. Whereas male and female subjects were almost equally distributed over the two conversational settings, almost twice as much female speech was recorded in the telephone dialogues. The numbers of subjects in Table 1 represent unique speakers, but speakers could participate in both multiple face-to-face and telephone conversations. Speakers who participated in both face-to-face and telephone conversations are treated as different speakers for each speech style.

Table 1: Overview of the material

	Male		Female	
	FTF	TD	FTF	TD
#unique subjects	50	40	58	61
duration(hh:mm:ss)	3:16:20	4:15:37	3:15:21	8:14:14
#utterances	9228	11557	9242	21290
#words/utterance	5.8	6.0	5.7	6.2
#words	53758	68937	52193	131994
#phonemes	152284	188438	148854	363942

3. Results and discussion

3.1. Speech rate

Speech rate is one of the most global measures describing the use of spoken language. Speech rate can be defined in many different ways, depending on whether the focus is on information transfer (normally expressed in terms of the number of words per second) or on the number of events per time unit (typically expressed in terms of the number of syllables or phonemes per second). Other variables determining the definition of speech rate are the inclusion or exclusion of silent pauses and the representation (orthographic or phonetic transcriptions) of the events under investigation. In order to analyze the average speech rate of the male and female subjects in the corpus, we first computed their *speaking rate* in terms of the number of words per second. Silent pauses and filled pauses were included in this measure. Subsequently, the *articulation rate* was computed in terms of the number of words per second, exclusive of silent pauses. The availability of manually checked phonetic transcriptions enabled us to compute the articulation rate in yet another way, namely in terms of the number of actually *realized* phonemes per second (again excluding silent pauses). The results of these measurements are presented in Table 2.

Table 2: Speech rate of male and female speakers (standard deviation in parentheses, significance indicated with *)

	Male	Female
words/s	3.72 (.49)	3.67 (.47)
words/s no pause	4.57 (.50) *	4.43 (.44) *
phonemes/s no pause	12.62 (1.32)	12.35 (1.23)

A *t*-test for independent samples on the speaking rate of our male and female subjects showed that the difference between 3.72 and 3.67 words/s is not significant ($p > .05$). This result is in line with [4], who did not find significant differences in speaking rate between male and female speakers in the American English Switchboard corpus. Assuming that [4] included pause durations in the measure for speaking rate, there seems to be a striking similarity between the speaking rate in Dutch and American English conversations (3.72 and 3.73 words/s for the male speakers; 3.67 and 3.68 words/s for the female subjects). This is probably related to the similarity between the word lengths in the two languages [7].

A *t*-test of the articulation rate, measured in words/s without silent pauses, showed that the male subjects produced significantly more words per second than the female subjects ($p=.028$). This result is in line with [8], who also found a significantly higher articulation rate for male speakers, measured in (orthographically based) syllables per second, pauses excluded. Relating the articulation rate to the results found on speaking rate suggests that the male speakers used more or longer silent pauses than the female speakers.

A statistical analysis of the articulation rate as measured in phonemes/s without silent pauses, in turn, did not reveal significant differences between male and female speakers. However, the two measured articulation rates allow us to conclude that the male subjects uttered more words containing fewer phonemes per second than the female speakers.

3.2. Disfluencies

Two types of disfluencies were studied: filled pauses (hereafter referred to as *fillers*, see [9]) and repetitions. For the purpose of our study, fillers were defined as those utterances in our data that were transcribed as “uh” and “uhm”. Repetitions were defined as repeated single orthographic words that have the same POS tag, e.g. “de de de uh de deur...” (“the the the uh the door”) – this utterance has two uninterrupted repetitions of the definite article “de”, hence, the repetition’s length is defined as three words. Words transcribed as “mm-hu” (“uh-huh”) and repetitions of the words “ja” and “nee” (“yes” and “no”) were excluded from our study since they are typically conventional backchannel signals rather than disfluencies [10].

Table 3 presents the statistics on the fillers and the repetitions. The percentages for fillers and repetitions were calculated relative to the total number of words uttered by males or females in face-to-face or telephone dialogues. The dialogue partner-dependent figures (MM (male-male), MF (male-female), and FF (female-female)) were calculated relative to the total number of words uttered by males or females in that specific dialogue context. The division of the dialogues in these three categories was straightforward for the telephone dialogues and for most of the face-to-face conversations. Ten out of 53 face-to-face conversations, however, were *multilogues* involving at least three speakers. These multilogues were categorized according to the gender that was dominant in terms of the number of speakers and the number of words.

Table 3: Statistics on fillers and repetitions.

	Male		Female	
	FTF	TD	FTF	TD
%fillers	3.47	4.59	2.38	3.42
%fillers initial	1.02	1.40	0.72	1.12
%fillers final	2.45	3.19	1.65	2.30
%fillers MM	3.43	5.42	-	-
%fillers MF	3.34	5.23	2.26	3.75
%fillers FF	-	-	2.48	3.24
%repetitions	2.29	2.51	1.18	1.35
%rep initial	0.69	0.74	0.38	0.45
%rep final	1.60	1.77	0.80	0.90
%rep MM	2.77	2.38	-	-
%rep MF	2.01	2.57	1.31	1.63
%rep FF	-	-	1.12	1.19

3.2.1. Fillers

Independent samples *t*-tests showed that the male speakers used more fillers than female speakers (3.47% and 4.59% vs. 2.38% and 3.42%, $p < .001$). Similar results were also reported by [5] in a study on American English telephone speech. Moreover, our data revealed that both male and female speakers used more fillers in telephone speech than in face-to-face conversations ($p < .001$). A closer look at the occurrence of the fillers showed that both male and female subjects used significantly more fillers after the third word of an utterance (depicted as ‘final’ in Table 3) rather than in the initial three words of an utterance (denoted as ‘initial’ in Table 3). Recall

that, according to the data in Table 1, the average length of an utterance was 5.9 words for both male and female speakers.

The more frequent use of fillers in telephone conversations (as opposed to conversations where the conversation partner was physically present), as well as the specific positioning of the fillers (more frequently in the final part of utterances than in the first three words) may indicate that these fillers have a *turn management* function [9].

The frequent occurrence of fillers in male speech and the data-derived hypothesis that fillers may have specific discourse functions, made us investigate whether the gender of the conversation partner influenced the use of fillers. Our results showed that fillers were significantly ($p < .001$) more used in conversations between two male subjects (3.43% and 5.42%) than in conversations between two female speakers (2.48% and 3.24%). Moreover, also in conversations with both male and female speakers, the male speech was characterized by a significantly higher use of fillers ($p < .000$) than the female speech 3.34% and 5.23% vs. 2.26% and 3.75%). These results are interesting in the light of [10], who studied the same data and reported significantly more overlapping speech (including fillers) in conversations between male speakers than in conversations between female speakers. As there is no reason to assume that male and female speakers deploy different planning strategies, the higher degree of overlapping speech in conversations with male speakers and their more explicit use of fillers towards the end of utterances suggest that fillers and overlapping speech may serve a discourse function rather than a planning function, as has long been assumed [9]. If this would be true, then male speakers can be considered to deploy more overt discourse management strategies than female speakers. Of course, the issue of turn management should be investigated in more detail.

3.2.2. Repetitions

Statistical analyses of the repetitions revealed that male speakers uttered significantly more repetitions than female speakers, regardless of the interlocutor's gender. No significant differences were found in the distribution of repetitions between face-to-face conversations and telephone dialogues. However, similar to the positioning of the fillers, male as well as female speakers used significantly more repetitions after the first three words of an utterance. This may indicate that in addition to fillers, also repetitions have a function in turn management.

Studying the possible influence of gender on the use of fillers and repetitions revealed interesting similarities and differences between the male and female speech. For one, both the male and female speakers used more fillers and repetitions in the final part than in the first three words of an utterance. Moreover, the male speech was characterized by a much higher frequency of fillers and repetitions, irrespective of the interlocutor's gender. These findings suggest that fillers as well as repetitions often act as tools for discourse management, and that male speakers are more outspoken than female speakers when it comes to discourse management. Again, a follow-up study would be worthwhile to investigate the hypotheses gathered from the data.

3.3. Word pronunciations

In order to get a first global indication of possible pronunciation differences between male and female speech, the manually verified phonetic transcriptions of our data were compared with a reference transcription. Since the manual production of a reference transcription is infeasible for large datasets, we resorted to a canonical reference transcription. This transcription was automatically obtained through a lexicon-lookup procedure in a lexicon comprising one canonical transcription for every word in the corpus.

First, the transcriptions were aligned with the reference transcription. Subsequently, the discrepancies between the transcriptions and the reference transcription were investigated and expressed in an overall disagreement measure defined as:

$$\%disagreement = \frac{Sub_{phone} + Del_{phone} + Ins_{phone}}{N_{phone}} * 100$$

i.e. the sum of all phone substitutions (Sub_{phone}), deletions (Del_{phone}) and insertions (Ins_{phone}) divided by the total number of phones in the reference transcription (N_{phone}).

The alignment and the computation of the percentage disagreement were performed with ADAPT [11], a dynamic programming algorithm that computes the optimal alignment between two strings of phonetic symbols according to a matrix in which the distances between phonetic symbols are defined in terms of articulatory features. Table 4 presents the global statistics of the alignment (for clarity, further breakdowns of the number of phone substitutions, deletions and insertions according to the gender of the conversation partners are omitted).

Table 4: Statistics on pronunciation differences

	Male		Female	
	FTF	TD	FTF	TD
%disagreement	16.74	16.82	16.45	16.79
%dis MM	16.22	16.92	-	-
%dis MF	17.23	16.76	15.84	16.46
%dis FF	-	-	16.85	16.99
%substitutions	8.00	8.55	7.67	8.23
%deletions	7.90	7.22	7.97	7.73
%insertions	0.84	1.05	0.81	0.83

Hardly any significant differences (independent samples t -tests, $p < .05$) were found between the overall disagreement and the total numbers of phone substitutions, deletions and insertions in male and female speech. Also the interlocutor's gender did not have a significant impact on the numbers. This is contradictory to the stereotypical idea that women articulate better than men [12].

Our global disagreement metric only provided valuable information on the similarities between the male and female pronunciation (measured in terms of the disagreement from a reference transcription). However, it may very well be that gender-specific differences can be found within the categories of phone substitutions, deletions and insertions. This hypothesis, however, is not supported by our initial study, and can therefore only be verified through a more thorough investigation of the data.

3.4. Use of words

Finally, we compared the male and female speakers' use of two frequently used parts of speech, viz. nouns and pronouns. The figures in Table 5 are calculated in proportion to the total number of words produced by males and females in the different speech styles.

No significant gender-specific differences were found in the occurrence of nouns. Interestingly, the ten most frequently used nouns did not differ between males and females. Most probably, this is due to the fact that most of the conversations were about daily activities ("beetje" ("bit"), "keer" ("time"/"turn"), "jaar" ("year"), "mensen" ("people"), "man" ("man"), "soort" ("kind"/"sort"), "dingen" ("things"), "week" ("week"), "uur" ("hour") and "tijd" ("time")).

An analysis of the parts of speech in our data showed that pronouns accounted for about 20% of all words, and that almost half of the pronouns, viz. the personal, possessive and reflexive pronouns, were used to relate to personal issues. Our data showed that the female speech in the telephone dialogues contained significantly ($p < .05$) more of these pronouns than the male speech, irrespective of the conversation partner. This confirms the assumed stereotypical difference between men and women, stating that women speak more often about personal issues than men do.

Table 5: Statistics on the use of nouns and pronouns

	Male		Female	
	FTF	TD	FTF	TD
%nouns	9.88	8.71	9.63	8.28
%nouns MM	10.28	9.58	-	-
%nouns MF	9.52	8.33	10.04	8.36
%nouns FF	-	-	9.33	8.23
%selected pronouns	7.74	6.97	7.88	8.06
%selected pron MM	8.15	6.44	-	-
%selected pron MF	7.53	7.22	8.04	8.08
%selected pron FF	-	-	7.76	8.06

4. Conclusion

The primary aim of this study was to investigate the possible interaction of a speaker's gender and his or her everyday speech and language. Simultaneously, we wanted to investigate the power of data-mining as an approach to get swift access to large data collections.

As to the first aim, interesting differences were found between male and female speech. First, we found the male speakers to have a higher articulation rate than the female speakers. Although this result is in line with [8], the measures for articulation rate deployed in [8] (orthographic syllables per second) and in our study (words per second) differ. The impact of these different measures on the results remains to be investigated. Second, our results showed that male speakers tend to use more or longer silent pauses and utter more words with fewer phonemes than female speakers. Third, we found higher proportions of fillers and disfluencies in male speech. However, both in male and female speech, fillers and repetitions are preferably positioned at the end rather than in the first three words of an utterance. This finding raises interesting questions about the relation between disfluencies and planning at the level of the utterance and the discourse. If utterance-final disfluencies are indeed used to signal that a

speaker intends to keep or cede the turn, they are probably less of an indication that the planning of the ongoing utterance is not yet complete. In this case, fillers and disfluencies can be regarded as mechanisms used for turn management. Evidently, a more detailed analysis of the data is indispensable to verify these hypotheses.

As to the second aim of our study, our results proved data-mining to be a useful method to explore large spoken language corpora. The increasing availability of large annotated spoken language corpora and the ease with which these corpora can be accessed make the data-mining approach likely to gain popularity in the future. Thanks to corpus-based research it is now possible to complement our knowledge of spoken language processes in real-life situations. This knowledge is extremely important for the development of applications and services that deal with real-life speech rather than intensively researched laboratory speech.

5. Acknowledgements

The authors thank Bram Elffers for writing ADAPT.

6. References

- [1] Godfrey, J.J., Holliman, E.C., & McDaniel J., "SWITCHBOARD: Telephone speech corpus for research and development", *Proceedings of ICASSP, San Francisco*, pp. 517-520, 1992.
- [2] Oostdijk, N., "The design of the Spoken Dutch Corpus", Peters P., Collins P., Smith A. (Eds) *New Frontiers of Corpus Research*, Rodopi, Amsterdam, pp. 105-112, 2002.
- [3] Furui, S., Maekawa, K., & Isahara, H., "A Japanese national project on spontaneous speech corpus and processing technology", *Proceedings of ISCA ITRW ASR2000, Paris*, pp. 244-248, 2000.
- [4] Syrdal A., "Acoustic variability in spontaneous conversational speech of American English talkers", *Proceedings ICSLP, Philadelphia*, pp. 438-441, 1996.
- [5] Schriberg, E., "To 'err' is human: ecology and acoustics of speech disfluencies", *Journal of the International Phonetic Association*, 31/1, pp. 153-169, 2001.
- [6] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D., "Effects of disfluencies, predictability, and utterance position on word form variation in English conversation", *J. Acoust. Soc. Amer.*, Vol. 113(2), pp. 1001-1024, 2003.
- [7] Baayen, R.H., Piepenbrock, R., & Gulikers, L. "The CELEX Lexical Database (Release 2)" [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania, 1995.
- [8] Verhoeven, J., De Pauw, G., & Kloots, H., "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands", *Language and Speech*, 47 (3), pp. 297-308, 2004.
- [9] Clark, H. & Fox Tree J., "Using *uh* and *um* in spontaneous speaking", *Cognition*, 84, pp. 73-111, 2002.
- [10] Ten Bosch, L., Oostdijk, N., & de Ruyter, J.P., "Turn-taking in social dialogues: temporal, formal and functional aspects", *Proceedings SPECOM, St. Petersburg*, CD-ROM, 2004.
- [11] Elffers, B. "ADAPT: Algorithm for Dynamic Alignment of Phonetic Transcriptions", *manual available online from <http://lands.let.ru.nl/literature/elffers.2005.1.pdf>*, 2005.
- [12] Trudgill, P. "Sex, covert prestige and linguistic change in the urban British English of Norwich", *Language in Society*, 1, pp. 179-195, 1972.