

The influence of social distance on speech behavior

Formality variation in casual speech

Kim Koppen¹, Mirjam Ernestus^{1,2} and Margot van Mulken¹

¹Radboud University Nijmegen, the Netherlands

²Max Planck Institute for Psycholinguistics, the Netherlands

Abstract: An important dimension of linguistic variation is formality. This study investigates the role of social distance between interlocutors. Twenty-five native Dutch speakers retold eight short films to confederates, who acted either formally or informally. Speakers were familiarized with the informal confederates, whereas the formal confederates remained strangers. Results show that the two types of interlocutors elicited different versions of the same stories. Formal interlocutors (i.e. large social distance) elicited lower articulation rates, and more nouns and prepositions, both indicators of explicit information. Speakers addressing the informal interlocutors, to whom social distance was small, however, provided more explicit information with an involved character (i.e. adjectives with subjective meanings). They also used the word *and* more often as a gap filler or as a way to keep the floor. Furthermore, a small social distance elicited more laughter, interjections, first-person pronouns and direct speech, which are all indicators of involvement, empathy and subjectivity.

Keywords: social distance, formality, audience design, language, speech behavior

1 Introduction

A large part of our daily life exists of interactions with our family, friends, neighbors, colleagues, officials and others. These interactions differ in the way interlocutors use language to express themselves, for instance in word choice, pronunciation and/or sentence structure (e.g. Firth 1968[1952-1959]; Halliday 1978; Reid 1956). One of the most important dimensions of this linguistic variation is formality (e.g. Biber and Conrad 2009; Biber et al. 1998; Heylighen and Dewaele 2002; Labov 2006 [1966]), but while most people can make an intuitive distinction between formal and informal manners of speech (Creber and Giles 1983; Lahiri 2011), it is an ongoing challenge to grasp the exact relation between particular speech situations and the corresponding linguistic characteristics. As Dittmar (2010) concludes, there is an urgent need to refine existing models of linguistic variation by investigating this relation in more detail. How exactly do speakers adapt their speech behavior according to the formality of the speech situation?

Various studies revealed that certain (para)linguistic features occur more in formal than in informal language or vice versa. The concept of formality is not as straightforward as one might think though, because formality is influenced by many parameters (Berruto 2010; Heylighen and Dewaele 2002), such as the setting, the topic of conversation, the modality (written versus spoken language), and the speaker's audience, which potentially have different impacts on speech behavior (Figure 1). Following Bell (1984, 2001: 143), who argued that "speakers design their style primarily for and in response to their audience", this paper focuses on the parameter audience and addresses the question whether speakers express the same idea differently to interlocutors with whom their relation is either formal or informal.

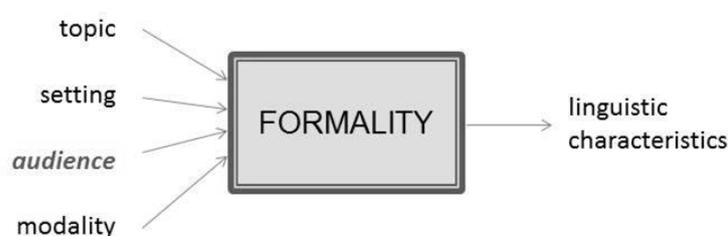


Figure 1. Parameters influencing degree of formality (adapted from Heylighen and Dewaele [2002]).

1.1 Formal and informal interlocutors

The way we speak has social implications (Richards and Schmidt 2010). Speaking too formally to your friends or addressing the mayor of your city too informally, even in a chance encounter, may lead to awkward situations. The ability to make the appropriate linguistic choices in audience design (Bell 1984, 2001) is thus an important skill that takes place "in terms of a wide range of linguistic-prosodic-nonverbal features" (Giles et al. 1991:7). Most (native) speakers learned to master this skill quite well, which, according to the Communication Accommodation Theory (CAT), allows speakers to

regulate the social distance between them and their audience (Giles and Powesland 1975; Gasiorek et al. 2015).

How formal or informal a speaker believes this audience to be, can be accounted for by the Language Expectancy Theory (LET: Burgoon et al. 2002). LET argues that people develop certain expectations and norms during lifetime that shape the way they address their audience. These expectations are raised by the addressee's individual characteristics, including appearance, personality and social status, as well as by relational factors such as how similar and equal in status interlocutors are and how well they know each other. This implicates that people do not only adapt their speech behavior to their interlocutor's speech, but additionally, take social characteristics like age, education, profession and even clothing into consideration, when they assess the formality of the speech situation.

1.2 Formal and informal speech behavior

Labov (2006 [1966]), Stolarski (2013) and Ernestus et al. (2015), among others, studied how speakers address their audience in formal versus informal settings or when discussing formal versus informal topics of conversation. Kouwenhoven et al. (2015) investigated informal versus formal speech as well, and examined speech behavior in non-native interaction. Here too, the subject of conversation was one of the main parameters of variation between the formal and informal condition. Biber (1988) and Biber et al. (1998) compared written to spoken language on several dimensions, among which formality. As in most of the other research, the content of the language in the compared modalities varied. These previous studies made important contributions to our understanding of formality, but they investigated other parameters than social distance and seldom the topic of conversation was equal in the formal and informal situations. The current study aims to complement the existing knowledge by investigating formality variation as a result of social distance, while keeping constant other parameters that could possibly influence the formality of the speech situation, being the topic of conversation, the setting and modality (Figure 1).

Several linguistic variables have been shown to differ as a function of formality, and the current study will investigate similar characteristics of speech. Heylighen and Dewaele (2002), for example, argue that relatively high frequencies of nouns, articles, adjectives and prepositions are associated with formal language, because speakers formulate their messages more explicitly and precisely in formal speech situations. Compare for instance the sentence *And then I saw this pretty blue bird that was singing beautifully!* with its more formal counterpart *In the willow at the lake was a blue, beautifully singing kingfisher.* The second sentence obviously contains more explicit information and therefore avoids ambiguities that could possibly emerge if only the information in the first sentence were provided. Informal language, on the other hand, the type of speech used among friends and family, is said to be characterized by more involved speech behavior with relatively high frequencies of interjections, verbs, adverbs and pronouns (Heylighen and Dewaele 2002) as can be seen in the example above. The first-person pronoun singular, proved to be especially useful as a measure of subjective speech behavior (Vis 2011). Its use reflects a person's ability to feel empathy, which is a characteristic of informal language (Scheibman 2002).

Differences in formal and informal speech behavior for other kinds of linguistic features than part of speech have been observed as well. For instance, the amount of laughter has been shown to vary between formal and informal speech situations (e.g. Garcia 2013). It is part of universal human vocabulary and has an important social function (Provine 1996). In most situations it is an indicator of informal speech behavior, although laughter can be a sign of nervousness or embarrassment as well. Furthermore, previous research detected that in formal language word type/token ratios are higher, longer words are used and sentence structures are more complex, especially in (formal) written versus (more informal) spoken language (Biber 1988; Biber et al. 1998).

The formality of the situation can also influence the flow of speech. When assessing the casualness of the speech in their corpus, Torreira, Adda-Dekker and Ernestus (2010) showed that disfluencies (hesitations, false starts and repetitions) occur more often in spontaneous speech than in careful (more formal) speech, which is in line with earlier findings (e.g. Shriberg 2001). On the contrary, it is known that people can get nervous addressing interlocutors in formal situations and as a result may make more speech errors (Carroll 1986). One of the few studies that specifically investigated the effect of social distance revealed that people tend to speak more slowly when they address strangers than friends or family members (Yuan, Liberman and Cieri 2006).

These studies show that variability in speech behavior is driven by differences in formality, and, importantly, that speakers adapt their speech behavior in various ways. However, only few of these studies specifically looked at social distance, and kept content a constant factor.

Heylighen and Dewaele (2002: 4-7) argue that a cluster of variables, the number of nouns, articles, prepositions and adjectives, provide us with information about the level of explicitness. Likewise, other variables that have been shown to differ as a function of formality in the literature discussed above can be clustered as well. Complexity of language on a lexical level can be represented by word type/token ratio, word length and on sentence level by frequencies of conjunctions. Differences in speech rate and numbers of hesitations and speech errors can be seen as indicators for adaptations speakers make in flow of speech. Last, laughter, the number of interjections, adverbs and first-person pronouns singular can give us insight about the degree of involvement (e.g. Biber et al. 1998; Garcia 1993; Heylighen and Dewaele 2002; Vis 2011).

In the present study, these clusters were investigated systematically and in relation to each other, and we controlled for other parameters influencing the formality of the speech situation, in particular for the topic of conversation (Figure 1). Table 1 provides an overview of the (para)linguistic features that were analyzed for the level of complexity, flow of speech, the level of explicitness and degree of involvement. For flow of speech, pause duration was incorporated as well, because it is closely related to speech rate (e.g. Bosker, Pinget, Quené, Sanders and de Jong 2013). The number of articles was not included in the analysis, because the occurrence of articles correlates with the occurrence of nouns. Furthermore, conjunctions were split in coordinating and subordinating conjunctions, because they incorporate different information about the nature of the complexity of the sentence. Based on previous findings, as discussed above, we expect language to be more involved when social distance is small, and more complex and explicit when social distance is larger, and that speakers also pay extra attention to their flow of speech when addressing formal interlocutors.

Table 1. Variables used to measure flow of speech, degree of involvement, the level of complexity and explicitness of information. Variables were measured in frequencies of occurrence, except for word length, word type/token ratio and articulation rate. For pauses, their frequency as well as their duration were measured.

Level of Complexity	Flow of Speech	Level of Explicitness	Degree of Involvement
Word length	Articulation rate	Nouns	Laughter
Word type/token	Pauses	Prepositions	First-person singular
Conjunctions	Hesitations	Adjectives	Interjections
- coordinating	Speech errors		Adverbs
- subordinating			

1.3 Formality and gender

Gender is a personal characteristic that plays an important role in society (e.g. Eckert 1989, 1990) and several (para)linguistic features mentioned above, indicating the degree of formality, are also found to differ between men and women. For example, word length varies with the degree of formality (e.g. Biber et al. 1998), but across gender as well (e.g. Newman, Groom, Handelman and Pennebaker 2008). The same holds for speech rate: Verhoeven, de Pauw and Kloots (2004) found a difference between men and women, and Yuan and colleagues (2006) found that speech rate varied with the degree of formality.

This raises the question, if and how gender, formality/social distance and speech behavior are related to each other. In addition to our main research question, we investigated whether variation in formal and informal language interacts with gender, which was also inspired by Dewaele’s statement (1998: 3): “If we look for differences between women’s speech and men’s speech, we need to be aware that other variables may intervene and bias the results.”

Earlier research produced contradictory results, for instance, regarding the use of first-person pronouns singular by men and women, as was observed by Newman et al. (2008). These kinds of conflicting findings might be explained when the influence of formality is taken into account. Dewaele (1998) indeed found a relation between formality and gender: women’s and men’s differed in informal chats, whereas their speech did not differ significantly during the more formal oral exams. Dewaele concluded that women adopt a more involved manner of speech in informal settings than men do. Eckert (1989) and Labov (1990) incorporated social class as an additional variable in their studies of formality (casual versus careful versus read speech) and showed that in the higher ranks of the socio-economic hierarchy gender differentiation is greater. This could mean that men’s and women’s speech deviate more when they address people with a higher social status than when they address people equal in status (small social distance). With these findings in mind, we investigated for all variables how men and women adapt their speech behavior to the formality of the situation.

1.4 Manipulation of formality

A formality corpus was created specifically for the purpose of this research. Each speaker in this corpus was provided with the same content in the form of short films. They were instructed to retell each film to the next participant and were told that this person had to retell the story again to yet another participant. In reality, the next participant was a confederate with a specific social status. Each speaker had to retell half of the films to a confederate who was instructed to act as a formal listener and the other half of the films to a confederate acting as an informal listener. Constructing the corpus in this way had the advantage that both speakers and stories in the formal and the informal part of our corpus are the same. As a consequence, we could compare descriptions of the same film scenes in both parts of our corpus directly.

We operationalized formality by selecting the confederates based on their general appearance, age, education and profession (cf. LET: Burgoon et al. 2002). For the informal confederates, we chose people similar to our participants, who were Dutch undergraduate students. The formal confederates, on the contrary, were higher in social status than our participants. Furthermore, clothing was used as a means to in- and decrease social distance between participant and confederate, because it has a major impact on their appearance and perceived social status (Slepian et al. 2015).

Importantly, the formal confederates remained strangers to the participants up until the actual recordings, since social status is known to become less important when people become more familiar with each other (Bell 1984; Douglas-Cowie 1978). For exactly the same reason, the participants did meet the informal confederates before the recordings. Because of the similarities between them, we expected the participants could easily identify themselves with these confederates and got acquainted with each other informally (Mitchell 1998).

Since this research is about the way people address other people they encounter in everyday life, we made sure the situation as a whole was not too formal. The formality continuum is a scale on which speech situations can be arranged according to their degree of formality (Ager 1990; Sanders 1993). As shown in Figure 1, various parameters determine how formal or informal a speech situation is and, as a result, where a situation is located on this continuum. The setting and the topics of the films were chosen in such way that our corpus positions itself on the informal side of the formality continuum, as depicted in Figure 2, with C_f and C_{inf} being respectively the formal and informal part of our corpus.



Figure 2. The area on the formality continuum our research focuses on. C_f and C_{inf} are respectively the formal and informal part of our corpus.

2 The formality corpus

Our formality corpus contains five hours of speech of twenty-five native Dutch speakers retelling short films to confederates, who were instructed to act as either formal or informal listeners. In this section, the creation of this corpus is described in detail: the key players in it, the recordings and orthographic transcription.

2.1 Participants and confederates

Twelve male and thirteen female students from Radboud University participated in the recordings. All participants were Dutch native speakers, aged between eighteen and twenty-seven years, with a mean of twenty-one years and reported no hearing, speech or language impairments.

To elicit either formal or informal speech behavior from the participants, we recruited two types of confederates. The formal confederates, one female and one male, were aged between fifty and sixty-five years. Both were highly educated (respectively a biology teacher and a philosopher) and had side jobs as surveillance guards at student exams at Radboud University. Two male and two female student assistants at Radboud University, aged between eighteen and twenty-four years, acted as informal listeners. Since the participants were also students, we expected they would easily identify themselves with these informal confederates. Both formal and informal confederates' genders were matched with the participants' genders.

In order to check whether our estimation of the confederates' general formality was agreed upon by others, five students, other than the participants in the experiment, scored the confederates' photographs. The formal confederates scored a minimum of four and the informal confederates a maximum of two on a five point formality scale, in which one stood for very informal and five for very formal.

The formal confederates were instructed to act like formal persons with authority and listen attentively to the participants with a neutral expression. To prevent other adaptation processes to interfere, the only interaction allowed was an occasional nod or interjection. This was practiced in a trail session. Attention was particularly paid to their clothing. This supposedly helped the confederates to play their role convincingly, because the way a person dresses does not only influence other persons (cf. our participants), but that person him- or herself too (Hannover and Kühnen 2002; Slepian et al. 2015). The formal male confederate was dressed in a dark suit with tie and the female confederate wore a black skirt with a beige jacket. The informal confederates, on the contrary, received instructions to dress informally, listen attentively with an inviting attitude, not to speak either, but smile, laugh and nod occasionally.

2.2 Films

Eight short films provided the speaking material, each with a duration of approximately two minutes. The films originated from various (short) film festivals and YouTube. Table 2 presents an overview of the films. Films were shown without sound and subtitles to prevent influencing the participant's speech behavior. The eight films formed four pairs, each pair representing a different genre. Participants retold one film of each pair to the

formal confederate and the other, similar film to the informal confederate. Films were distributed in such way that each film was retold equally often to the two types of confederates. To avoid sequence effects, the order in which the films were shown was randomized for each participant. Half of the participants started with retelling films to the formal confederate and the other half to the informal confederate.

Table 2. Films providing the speaking material.

Genre	No.	Title (Director[s] year)	Description
Cartoon	1a	Lopoo & Donkey (Liu 2009)	A man and a donkey are lost in the desert without food or drinks. Then they find a tree with only one apple. Does true friendship survive?
	1b	Octapodi (Bocabeille et al. 2007)	Two octopi heavily in love are fighting for their lives with a stubborn restaurant cook in a comical escape through the streets of a Greek village. Does true love conquer?
Instruction	2a	Pop-up Tent (Quechua 2011)	How to pitch and fold a pop-up tent, filmed in a campsite-like surrounding.
	2b	Magic Revealed (Milleaccendini 2012)	How to perform a magic trick with matches.
Story	3a	Coasting (Forcolini 2011)	Two old friends set off on a countryside adventure, searching for the best that outdoor offers. Will they be in time to catch a glimpse of the rare specimens they're looking for?
	3b	The Black Hole (Sansom and Williams 2008)	A sleep deprived office worker photocopies a "Black Hole" late one night that enables him to pass through everything. Suddenly the possibilities seem endless. Will greed get the better of him?
Demonstration	4a	Parrot Tricks (Sazhin 2010)	A Senegal parrot performs twenty parrot tricks in just two minutes.
	4b	Rabbit Tricks (Orr and Lewin 2006)	Rabbits perform amazing tricks, starring bunnies rescued from shelters.

2.3 Procedure

A meeting between the informal confederate and the participant took place just before the recordings. When the participant arrived, the experimenter announced that the experiment would be delayed and invited the participant to wait together with another participant, who actually was the informal confederate with the assignment to become acquainted informally. After approximately five minutes they were informed that the experiment would start.

In case of the order formal - informal, the recording procedure was as follows. Both the informal confederate and the participant were brought to the experiment room. They were told they were participating in an experiment called "Retell the Story" and one (the real participant) had to retell several short films to the other, who would have to retell the stories to yet another participant (which would actually never happen). The participant watched the films in a small sound attenuated cabin with a head-mounted microphone. After each film, the door was opened and the participant retold the film to the confederate sitting in front of the cabin. Audio and video recordings were made, but the participant

was kept unaware of the video recording to keep the situation as informal as possible. Therefore, the camera was not directed at the participant, but was slightly offset. Additionally, the camera was surrounded by related unused equipment and an unplugged cable hung from the camera. After watching and retelling the films (one by one) to the informal confederate, the participant had a short break in which the experimenter took the confederate to another room as if (s)he were to retell the stories to the following participant. The formal confederate was then brought in for the second session and introduced him- or herself with their surname. The experimenter addressed this confederate formally with the Dutch formal second-person pronoun *U* and avoided any laughing and chit chat in order to reinforce the confederate's social status and the experimenter asked for permission to use the video recordings. Furthermore, the experimenter told the participant the procedure would be the same as in the previous session.

The procedure for the order formal - informal was slightly different. Only the participant was brought to the experiment room and the informal confederate was asked to wait while the formal confederate participated in the first session. During the break between the two sessions, the informal confederate and the participant spoke briefly to each other to re-establish their informal connection before proceeding.

2.4 Perception of formality

After the recordings, all participants and confederates completed a questionnaire: both scored the degree of formality displayed by the other. Moreover, the participants scored to what degree the confederates influenced the way they described the film.

The participants' questionnaires were analyzed with Classification and Regression Tree analysis (CART). A cost-complexity pruning algorithm was applied to increase prediction accuracy (Baayen 2008). First, participants rated the confederates' formality on a six-point Likert scale ranging from *very informal* (1) to *very formal* (6). The formal confederates were perceived significantly more formal ($M = 5.1$) than the informal confederates ($M = 2.0$) on average. The confederates' specific characteristics proved to be relevant for the participants' perception of formality as well; the formality ratings for facial expression ($M_{\text{formal}} = 5.2$, $M_{\text{informal}} = 2.1$), body language ($M_{\text{formal}} = 5.1$, $M_{\text{informal}} = 2.0$), clothing ($M_{\text{formal}} = 5.3$, $M_{\text{informal}} = 2.3$), age ($M_{\text{formal}} = 5.0$, $M_{\text{informal}} = 2.4$) and hairdo ($M_{\text{formal}} = 4.6$, $M_{\text{informal}} = 2.5$) were all significantly different between the two types of confederates. Participants did not rate female and male confederates significantly differently. We can conclude that, according to the participants, the confederates' appearance and behavior were in line with our intentions.

Furthermore, participants were asked to what extent the confederates influenced their speech behavior. For this question, the scale ranged from *no influence* (1) to *great influence* (6). Participants felt moderately to strongly influenced: pronunciation ($M = 3.6$), detail of the story ($M = 3.4$), word choice ($M = 3.7$), facial expression ($M = 3.9$), body language ($M = 4.0$), intonation ($M = 4.2$) and speech rate ($M = 4.3$).

Although the confederates were aware of the experiment's purpose, we still asked them to rate how formal the participants' speech behavior was. Paired t-tests showed that the formal and informal confederates' ratings differed significantly for all parameters (all $ps < 0.001$): pronunciation ($M_{\text{formal}} = 4.1$, $M_{\text{informal}} = 2.2$), word choice ($M_{\text{formal}} = 4.2$, $M_{\text{informal}} = 2.3$), facial expression ($M_{\text{formal}} = 4.4$, $M_{\text{informal}} = 2.0$), body language ($M_{\text{formal}} = 4.4$, M_{informal}

= 2.1), intonation ($M_{\text{formal}} = 4.2$, $M_{\text{informal}} = 2.1$) and speech rate ($M_{\text{formal}} = 3.8$, $M_{\text{informal}} = 2.2$). No gender effects emerged.

Based on these results, we can safely assume that our formality manipulation succeeded. The results indicate that the participants were influenced by the confederates and felt they adapted their speech accordingly. The participants' speech behavior was also perceived as more formal by the formal confederates than by the informal confederates.

2.5 Annotation

The corpus consists of 100 informal retellings with a total duration of 143 minutes and 27.580 words and 99 formal retellings with a total duration of 149 minutes and 26.636 words (one formal retelling is missing due to a technical problem). Native speakers of Dutch orthographically transcribed the corpus in Praat, a computer program for analyzing speech (Boersma and Weenink 2012). Boundaries were put around stretches of uninterrupted speech, called chunks, each having a duration of approximately 3 seconds. Speech errors, self-corrections, hesitations, broken words, repetitions, onomatopoeia and neologisms were labelled. Laughter, pauses, breaths and other non-speech sounds were also coded and put in separate chunks.

FROG, a Dutch morpho-syntactic analyzer and dependency parser (Van den Bosch et al. 2007) was used to make an automatic Part of Speech annotation. Due to the size of our corpus, it was only feasible to manually check whether the most frequently used words had been assigned the right word class considering the context in which they occurred. For the other words, we sorted Frog's output on part of speech and made a broad scan. The overall performance appeared to be good enough to rely on for further analysis.

3 Effects of social distance: results and discussion

This study investigated the effect of social distance on the (para)linguistic variables listed in Table 1, which function as indicators for complexity of language, flow of speech, level of explicitness and degree of involvement. Most of these indicators are frequency counts. In order to make fair comparisons between the formal and informal retellings, counts were normalized (e.g. Biber et al. 1998) to frequencies of occurrence per 100 words (i.e. a percentage of the total number of words in a retelling). All analyses were performed in R, a software environment for statistical computing (R Development Core Team 2015). Most of the variables were analyzed by means of linear mixed models, with formality and gender and their interaction as fixed predictors, and with speaker and film as crossed random factors, except when mentioned otherwise. First, we determined which fixed predictors were statistically significant by comparing the Akaike Information Criteria of the theoretically relevant models (AIC: Akaike 1973). Secondly, we tested whether the random intercepts and random slopes for formality contributed to the model by comparing AIC values of the models with and without random predictors. Holm-Bonferroni correction was applied to control for type I errors (Holm 1979). This procedure is as follows: the unadjusted p-values are ranked in decreasing order of

significance and then the α -level is adjusted sequentially for each p-value. Only p-values below their individually adjusted α -level are considered significant.

3.1 Level of complexity

Indicators for complexity were word length, word type/token ratio and frequencies of occurrence for subordinating and coordinating conjunctions. Table 3 provides an overview of the mean frequencies for these indicators, split by gender and formality. Word length was measured in number of syllables and is only based on content words, excluding false starts. The formal confederates did not elicit significantly longer words than the informal confederates, but the analysis revealed a significant effect of gender ($\beta = -0.09$, $t = 3.33$, $p < 0.005$): on average, male participants ($M = 1.69$, $SD = 0.14$) used slightly longer words than female participants ($M = 1.61$, $SD = 0.15$). Results were similar when word length was measured in the number of orthographic characters. The average word type/token ratio was about equal in the formal and informal part of the corpus (Table 3).

In contrast, the frequency of occurrence of coordinating conjunctions was significantly affected by both formality ($\beta = 0.41$, $t = 2.35$, $p < 0.05$) and gender ($\beta = 1.98$, $t = -4.06$, $p < 0.0005$), as can be seen in Table 3. Coordinating conjunctions occurred more frequently in the informal setting, when social distance was small, and female participants used more coordinating conjunctions compared to male participants. This may be contra intuitive, because one may expect more complex sentences in more formal settings. Inspection of the data taught us that this result is mainly due to the high incidence of the connector *and*. This word occurred mainly in sentence initial position, as exemplified by utterances [1a] from the informal part of the corpus versus [1b] from the formal part.

- (1) a. I (P23, f)¹ *En eerst deed hij het met de normale kop*
'And first he did it with the normal top'
- b. F (P20, m) *Eerst begint hij met het zwartmaken van de top*
'First he starts with blackening the top'

Given the position of *and* in the sentence, it seems questionable whether this result gives us insight in the complexity of sentence structure, because in most cases *and* does not have a coordinating function. Earlier research indeed showed that *and* is used not only to link at a syntactic/semantic level, but also at a pragmatic/discourse level. In her research on sentence-initial *and* Dorgeloh (2004) shows that "*and* possesses a natural propensity for functioning as an initiator of ongoing discourse". On a discourse level *and* thus functions to conjoin speech acts (Sweetser 1990: Chapter 4), rather than content items, being referred to as "exchange coordination" by Culpeper and Kytö (2000). We argue that, in our corpus, *and* mostly functions as a gap filler or as a way to keep the floor and should rather be associated with flow of speech as an indicator of informality.

¹ This code, identifying the participant and the formality condition, consists of the letter F for formal or I for informal and, between brackets, the participant's number and the letter f or m for the participants' gender.

addressing the informal confederate. According to Haeseryn and colleagues (1997), objective adjectives (such as *blue* and *rectangular*) qualify an object's concrete properties like color, shape or material, whereas the use of subjective adjectives (such as *nice* or *interesting*) relies on personal and intuitive criteria. Compare, for instance, the adjectives in utterances [3a] and [3b], which describe the same film scene, but were taken from respectively the informal and the formal parts of our corpus.

- (3) a. I (P24, m) *boven in de boom is een mooie rode glanzende appel*
 'up in the tree is a beautiful red shiny apple'
- b. F (P9, f) *een boom met een appel, een rode appel*
 'a tree with an apple, a red apple'

We therefore reanalyzed our data and classified adjectives as objective or subjective and incorporated the classification as a predictor in the analysis of the frequency of use of adjectives. The analysis revealed a highly significant interaction effect between formality and subjectivity ($\beta = 0.36, z = 3.92, p < 0.00001$) accompanied by a simple effect of subjectivity ($\beta = -1.20, z = -17.21, p < 0.00001$). The proportion of subjective adjectives was significantly lower than the proportion of objective adjectives in both the formal and the informal setting, but more importantly, indeed, subjective adjectives occurred more frequently when the participants retold the films to the informal confederates than to the formal confederates, whereas the number of objective adjectives was the same for both parts of our corpus (Figure 3). This additional analysis shows that, although subjective adjectives provide the listener with more explicit information, these adjectives also reflect involvement.

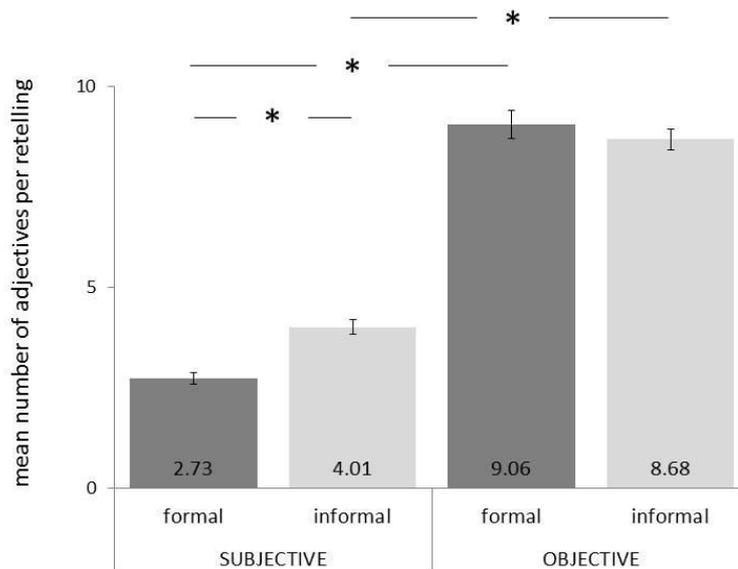


Figure 3. Number of subjective and objective adjectives in the formal and informal part of the corpus.

Table 3. Word length in average number of syllables per word, word type/token ratio and frequencies per word class in percentages split by formality and gender. Since not all word classes were analyzed, frequencies do not add up to 100%.

	FORMAL		INFORMAL	
	Male	Female	Male	Female
Word length	1.70	1.59	1.69	1.62
Type/Token ratio	44.46	43.29	45.63	42.44
Nouns	14.09	13.30	13.10	12.75
Prepositions	10.31	9.99	9.98	9.36
Adjectives	4.61	4.78	5.42	5.10
First-person pronouns	0.32	0.31	0.48	0.48
Adverbs	13.01	14.10	12.89	15.14
Interjections	1.58	1.80	2.12	2.37
Coordinating conjunctions	4.97	6.76	5.28	7.35
Subordinating conjunctions	1.10	1.22	0.98	1.09

3.4 Degree of involvement

Incidences of laughter and frequencies of occurrence for adverbs, interjections and first-person pronouns singular were used to measure the degree of involvement (Table 1).

Throughout our formality corpus, laughter was always the participant's initiative, because the confederates were only allowed to backchannel and behaved accordingly. There was laughter in about half of the retellings and when laughter was present, it occurred only once or twice. We therefore tested for the absence versus presence of laughter with a generalized linear mixed model with the binomial link function. This analysis showed effects of both formality ($\beta = 1.35, z = 3.25, p < 0.005$) and gender ($\beta = -3.25, z = -5.25, p < 0.001$): laughter was present more often in the informal than in the formal setting and female participants laughed more frequently than male participants.

Our finding that laughter must be regarded an indicator of informal language is, apart from earlier mentioned work on spoken language, also supported by research on expressions of laughter in written texts. Several studies (Drouin and Davis 2009; Grace et al. 2015) showed that the use of textisms (including smileys, LOL, haha etc.) is mainly reserved for causal text messages and considered to be inappropriate in formal writing. Furthermore, analogous to our (spoken) corpus, female writers use/report to use more emoticons (of which smiley faces being the most frequent) than men (e.g. Tossell et al. 2012; Rosen et al. 2010).

The results for adverbs ($\beta = 1.58, t = -2.23, p < 0.05$) shows that the female participants in our corpus were the more frequent users in comparison with the male participants (Table 3), but there was no formality effect.

Other distinctive features of formality in the cluster involvement appeared to be the frequencies of interjections ($\beta = 0.57, t = -2.23, p < 0.0001$) and first-person pronouns ($\beta = 1.08, t = 3.22, p < 0.005$) (Table 3). Participants addressing an informal listener were more likely to use those words. As can be seen in examples (4-6), interjections and constructions with *I 'ik* provide speakers with the means to make a story more vivid

3.5 Gender and formality

As mentioned in the Introduction, we investigated as well whether formality variation is different for male and female speakers, that is, whether the formality range for men and women is only different when they address informal interlocutors or, on the contrary, whether men’s and women’s speech deviate more when addressing formal interlocutors (i.e. when social distance is larger).

Interestingly, although laughter and higher frequencies of adverbs (both in the cluster involvement), and coordinating conjunctions (cluster flow of speech) are regarded as characteristics of informal speech behavior, women tended to use these (para)linguistic features more than men also when addressing the formal confederates. Men, on the other hand, used longer words (cluster complexity), usually associated with formal language, than women, not only when speaking to the formal, but as well when addressing the informal confederates. Thus, no interactions between gender and formality were found, only simple effects, which means that the influences of gender and formality on these linguistic variables were additive: differences between men’s and women’s speech were about as large in the formal part of the corpus as in the informal part of the corpus. These findings suggest that in casual speech, at least for these (para)linguistic variables, the whole formality range of men and women is slightly shifted relative to each other.

3.6 Summary of the results

Table 4 provides an overview of how social distance affected the participants’ speech behavior with respect to complexity of language, flow of speech, the level of explicitness and degree of involvement. Conversational partners to whom social distance was larger elicited a slightly lower articulation rate, and more explicit information, as evidenced by the higher frequency of nouns and prepositions, (Table 4: formal > informal), but not more complex language (Table 4: formal ≈ informal).

Table 4. Influences of social distance on indicators of complexity, flow of speech, level of explicitness and degree of involvement, reflecting formal and informal speech behavior. Indicators in italics were significantly different for gender. Indicators were measured in frequencies of occurrence, except word length (number of syllables per word, word type/token (ratio) and articulation rate (number of syllables per second)).

	Level of complexity	Flow of speech	Level of explicitness	Degree of involvement
Formal > Informal		Articulation rate	Nouns Prepositions	
Informal > Formal		<i>Coordinating conjunction “and”</i>	Adjectives (in total) Subjective adjectives	<i>Laughter</i> First-person singular Interjections Direct speech
Formal ≈ Informal	<i>Word length</i> Type/token ratio Subordinating conjunctions	Pausing Hesitations Speech errors	Objective adjectives	<i>Adverbs</i>

On the other hand, when social distance was small, speakers provided their interlocutor with more subjective adjectives, which are indicators of explicit information as well. Furthermore, the word *and* occurred more frequently, functioning as a gap filler or as a way to keep the floor. Last, speakers were more likely to show involved speech behavior when social distance was small: they laughed more often and also they made the story more vivid, gave more evaluative comments on the films and showed empathy by using more interjections, first-person pronouns singular and direct speech (Table 4: informal > formal).

4 General discussion

This study examined how social distance affects speech behavior and addressed the question whether speakers express the same information differently to formal interlocutors than to interlocutors with whom their relation is more informal. Twenty-five participants retold eight short films to confederates, who were instructed to act and dress either formally or informally. The confederates were selected based on their general appearance, age, profession and education, and their genders were matched with each participants' gender. Furthermore, the formal confederates were strangers to our participants, whereas the participants met the informal confederates just before the recordings for the corpus.

The data from our corpus show that just a difference in social distance between speakers is enough to affect speech behavior, despite the many other variables that determine the formality of a speech situation. This effect emerged in our study, even though we compared two instances of relatively casual speech in our formality corpus, instead of extremes on the formality scale: the interlocutors differed in age, clothing, behavior and in whether they had met the speaker before, but in both settings, they engaged in interactions that they may have on a daily basis. As a result, the speech in our corpus is fairly casual, meaning that we investigated variation in speech behavior at the informal side of the formality continuum (Figure 2).

Nevertheless, the effect of social distance was sufficiently large to be noticed by the speakers and their interlocutors themselves as well. The answers to our questionnaire indicate that speakers felt they had adapted their way of retelling the films to the degree of formality the listeners displayed. The participants' speech behavior was also perceived as more formal by the formal confederates than by the informal confederates.

Our analyses focused on four clusters of (para)linguistic variables that previous literature reported to vary as a function of formality (e.g. Biber et al. 1998; Garcia 2013; Heylighen and Dewaele 2002; Vis 2011). Several indicators of flow of speech, level of explicitness and degree of involvement showed clear effects of social distance, but this was not the case for indicators of linguistic complexity (see Table 4).

These previous studies, on which our expectations were based, however, either did not manipulate social distance, but one of the other parameters determining degree of formality (Figure 1), did not keep content a constant factor or did not focus on the informal part of formality scale, as the current study did. We therefore propose that the differences between our results and previous results highlight the unique effects of social distance on speech behavior in relatively informal situations. This implies that it is

important to distinguish between the different parameters influencing the degree of formality, such as the topic of conversation, the modality or the setting.

We believe that it is intuitively plausible that speakers adapt their speech behavior to the formality of the situation differently if social distance is manipulated instead of other indicators of formality. Reasoning from a CAT's perspective (Gaisorek et al. 2015; Pitts and Harwood 2015), a speaker's strong intention to enhance comprehension during formal presentations, for example, could explain that other research found a larger effect of formality on complexity or flow of speech because they varied *settings* (e.g. environments) rather than social distance between the speakers. It is only very likely that our participants would have used a different set of formality features if they had been requested to retell the stories as part of an assessment procedure for a job interview in a more formal setting than in the current experiment. This calls for future research that directly compares the effects of different indicators of formality.

Importantly, within the cluster involvement, almost all indicators increased in frequency when social distance was small. We found that the use of first-person pronouns, interjections was strongly related to the use of direct speech. Together, these variables indicate that speakers tended to make a story more vivid, more often choose to express their own thoughts/feelings, and empathize with other persons when talking to an informal interlocutor. This advocates for the incorporation of direct speech in the cluster of indicators of the formality of speech representing involvement.

This cluster also has to be extended with subjective adjectives. We found that speakers tended to use more subjective adjectives talking to an informal than to a formal interlocutor. We believe that the higher frequency of subjective adjectives reflects subjectivity and involvement. At the same time these adjectives are indicators of explicit information. This leads to the conclusion that involved language is not necessarily less explicit and less rich in information than formal speech behavior, but that social distance influences the type of information a speaker shares. This calls for rethinking the earlier presumed relation between involvement and explicitness (e.g. Biber et al. 1998): we argue that involvement and explicitness are not opposites of each other, but that they must be regarded as two, closely related, but individual dimensions of variation.

Furthermore, within the cluster explicitness, regarding frequencies of occurrence of nouns, we found indications that speakers omit nouns more often when they address interlocutors with whom their relation is informal, which would be worthwhile investigating further.

Another unexpected finding concerned the coordinating conjunction *and*. We expected this word to indicate the complexity of linguistic structure, but interestingly, this word was often used as a gap filler or as a way to keep the floor. As such, the frequency of occurrence of *and* is an indicator of the speaker's wish to produce continuously flowing connected speech (Culpeper and Kytö 2000; Dorgeloh 2004) even if this leads to the production of less (or non-) informative words. The coordinating conjunction *and* is thus more an indicator of informal than of formal speech behavior and occurs in situations when social distance is small.

This study focused on the effects of social distance on speech behavior. We showed how people use language talking to formal and informal interlocutors and that they share the same information in a different way. The next question is how listeners perceive these differences between formal and informal speech behavior. Asking listeners to evaluate excerpts of the same film scenes from both parts of the formality

corpus would provide insight in how they interpret the differences. It is also important to further investigate the relation between social distance and speech user characteristics, for example, whether men and women perceive differences in social distance differently or how age is related to formal and informal speech behavior. Furthermore, it would be interesting to examine differences between formal and informal non-verbal communicative behavior. Finally, in addition to this research in which we addressed how people adapt their speech behavior to equal versus higher status interlocutors, it would be worthwhile to investigate how higher status speakers adapt to people equal and lower in social status than themselves.

To conclude, this research has shown that social distance has a substantial influence on speech behavior: flow of speech, the level of explicitness and degree of involvement all vary depending on how formal or informal the relation between interlocutors is. Furthermore, this study indicates that the relation between the different speech variables and degree of formality is more complex than is typically assumed.

References

- Ager, Dennis. 1990. *Sociolinguistics and contemporary French*. Cambridge: Cambridge University Press.
- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (eds.), *Second international symposium on information theory*, 267–281. Budapest: Akademiai Kiado.
- Baayen, Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.
- Bell, Allan. 2001. Back in style: Reworking audience design. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 139–169. Cambridge: Cambridge University Press.
- Berruto, Gaetano. 2010. Identifying dimensions of linguistic variation in a language space. In P. Auer & J Schmidt (eds.), *Language and space: An international handbook of linguistic variation. Volume 1: Theories and methods*, 226–241. Berlin: Walter de Gruyter.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge University Press.
- Boersma, Paul & David Weenink. 2012. *Praat: Doing phonetics by computer* [Computer program]. Version 5.3.04. <http://www.praat.org/> (accessed 16 January 2012).
- Bosch, Antal van den, Bertjan Busser, Sander Canisius & Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. In P. Dirix, I. Schuurman, V. Vandeghinste & F. van Eynde (eds.), *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*, 99–114. Leuven: Belgium.
- Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H. De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175.
- Burgoon, Michael, Vickie Pauls Denning & Laura Roberts. 2002. Language expectancy theory. In James Price Dillard & Michael Pfau (eds.), *The persuasion handbook: Developments in theory and practice (Chapter 7)*, 117–136. Sherman Oaks, CA: Sage.
- Carroll, David. 1986. *Psychology of language*, 253–256. Pacific Grove, CA: Brooks/Cole Pub. Co.

- Creber, Clare & Howard Giles. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences* 5(2). 155–161.
- Culpeper, Jonathan & Merja Kyto. 2000. The conjunction and in early modern English: Frequencies and uses in speech-related writing and other texts. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg & C.B. McCully (eds.), *Generative theory and corpus studies: A dialogue from 10ICEHL*, 299–326. Berlin: Mouton de Gruyter.
- Dewaele, Jean-Marc. 1998. The effect of gender on the choice of speech style. *ITL Review of Applied Linguistics* 119–120. 9–25.
- Dittmar, Norbert. 2010. Register. In Mirjam Fried, Jan-Ola Östman & Jef Verschueren (eds.), *Variation and change: Pragmatic perspectives (Handbook of Pragmatics Highlights)*, 221–233. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Dorgeloh, Heidrun. 2004. Conjunction in sentence and discourse: Sentence initial and and discourse structure. *Journal of Pragmatics* 36. 1761–1779.
- Douglas Cowie, Ellen. 1978. Linguistic code-switching in a Northern Irish village: Social interaction and social ambition. In P. Trudgill (ed.), *Sociolinguistic patterns in British English*, 37–51. London: Edward Arnold.
- Drouin, Michelle & Claire Davis. 2009. R u txtng? Is the use of text speak hurting your literacy? *Journal of Literacy Research* 41. 46–67.
- Eckert, Penelope. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1. 1245–1267.
- Ernestus, Mirjam, Iris Hanique & Erik Verboom. 2015. The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics* 48. 60–75.
- Firth, John Rupert. 1968 [1952–1959]. *Selected papers of JR Firth*. Bloomington: Indiana University Press.
- Garcia, Angela Cora. 2013. *Understanding talk in formal and informal settings*. London: Bloomsbury.
- Gasiorek, Jessica, Howard Giles & Jordan Soliz. 2015. Accommodating new vistas. *Language and Communication* 41. 1–5.
- Giles, Howard, Justine Coupland & Nicolas Coupland. 1991. Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland & N. Coupland (eds.), *Contexts of accommodation*. New York, NY: Cambridge University Press.
- Giles, Howard & Peter F. Powesland. 1975. *Speech style and social evaluation*. Academic Press.
- Grace, Abbie, Nenagh Kemp, Frances H. Martin & Rauno Parrila. 2015. Undergraduates' attitudes to text messaging language use and intrusions of textisms into formal writing. *New Media & Society* 17(5). 792–809.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij & Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst* [2 banden]. Tweede, geheel herziene druk. Groningen: Martinus Nijhoff Uitgevers/Deurne: Wolters Plantyn.
- Halliday, M. A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Hannover, Bettina & Ulrich Kühnen. 2002. The clothing makes the self via knowledge activation. *Journal of Applied Social Psychology* 32. 2513–2525.
- Heylighen, Francis & Jean-Marc Dewaele. 2002. Variation in the contextuality of language: An empirical measure. *Foundations of Science* 7(3). 293–340.
- Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2). 65–70.
- Kouwenhoven, Huib, Mirjam Ernestus & Margot van Mulken. 2015. Register variation by Spanish users of English: The Nijmegen Corpus of Spanish English. *Corpus Linguistics and Linguistic Theory*. doi:10.1515/cllt-2013-0054.
- Labov. 1990. The intercourse of sex and social class in the course of linguistic change. *Language Variation and Change* 2. 205–254.

- Labov, William. 2006 [1966]. *The social stratification of English in New York City*, 2nd edn. Cambridge: Cambridge University Press.
- Lahiri, Shibamouli, Prasenjit Mitra & Xiaofei Lu. 2011. Informality judgment at sentence level and experiments with formality score. In Gelbukh A. (ed.), *Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science* 6609. 446–457. Heidelberg: Springer-Verlag Berlin.
- Mitchell, Mary with John Corr. 1998. *The first five minutes: How to make a great first impression in any business situation*. New York, NY: Wiley.
- Newman, Matthew L, Carla J. Groom, Lori D. Handelman & James W. Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* 45(3). 211–236.
- Pitts, Margaret J & Jake Harwood. 2015. Communication accommodation competence: The nature and nurture of accommodative resources across the lifespan. *Language & Communication* 41. 89–99.
- Provine, Robert R. 1996. Laughter. *American Scientist* 84(1). 38–45.
- R Development Core Team. 2015. R: *A language and environment for statistical computing* [Internet]. Vienna, Austria: R Foundation for Statistical Computing 2013. Document freely available on the internet at: <http://www.r-project.org>.
- Reid, Thomas B. W. 1956. Linguistics, structuralism and philology. *Archivum Linguisticum* 8(1). 28–37.
- Richards, Jack C. & Richard Schmidt. 2010. *Longman dictionary of language teaching and applied linguistics*. Harlow: Pearson Education Limited.
- Rosen, Larry D., Jennifer Chang, Lynne Erwin, L. Mark Carrier & Nancy A. Cheever. 2010. The relationship between “textisms” and formal and informal writing among young adults. *Communication Research* 37. 420–440.
- Sanders, Carol. 1993. *French today: Language in its social context*. Cambridge: Cambridge University Press.
- Scheibman, Joanne. 2002. *Point of view and grammar: Structural patterns of subjectivity in American English conversation* (Vol. 11). Amsterdam: John Benjamins Publishing.
- Shriberg, Elizabeth. 2001. To ‘errrr’ is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31(1). 153–169.
- Slepian, Michael L, Simon N. Ferber, Joshua M. Gold & Abraham M. Rutchick. 2015. The cognitive consequences of formal clothing. *Social Psychological and Personality Science* 6(6). 661–668.
- Stolarski, Lukasz. 2013. Style-shifting as a function of multiple factors: A corpus based study. *Token: A Journal of English Linguistics* 2. 245–263.
- Sweetser, Eve E. 1990. From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure. *Cambridge Studies in Linguistics* 54. Cambridge University Press.
- Torreira, Francisco, Martine Adda-Decker & Mirjam Ernestus. 2010. The Nijmegen corpus of casual French. *Speech Communication* 52(3). 201–212.
- Tossell, Chad C., Philip Kortum, Clayton Shepard, Laura H. Barg-Walkow, Ahmad Rahmati & Lin Zhong. 2012. A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior* 28. 659–663.
- Verhoeven, Jo, Guy de Pauw & Hanne Kloots. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3). 297–308.
- Vis, Kirsten. 2011. *Subjectivity in news discourse: A corpus linguistic analysis of informalization*. Oisterwijk: Uitgeverij BOXPress.
- Vliegen, Maurice. 2014. ‘Allemaal subjectief’ sneerde ze. Citaatuitleidende werkwoorden in landelijke Nederlandse dagbladen in 1950/1 en 2002. *Tijdschrift voor Taalbeheersing* 36(2). 197–224.

Yuan, Jiahong, Mark Liberman & Christopher Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. *Proceedings of Interspeech 2006*. 541–544.

References of the films

Bocabeille, Julien, Francois-Xavier Chanfoux, Olivier Delabarre, Thierry Marchand, Quentin Marmier & Edmud Mokhberí. 2007. *Octapodi*. Gobelins l'école de l'image.

Liu, Qianqian 2009. *Lopoo & Donkey*. Seven animation studio.

Forcolini, Federico. 2011. *Coasting*. Well Fed Films.

Sansom Philip, Olly Williams. 2008. *The Black Hole*. HSI Films.

Sazhin, Michael. 2010. *20 Parrot Tricks in 2 Minutes - Kili Senegal Parrot*.

Orr, Joan and Teresa Lewin. 2006. *Amazing Rabbit Tricks*. Clickerbunny.

Milleaccendini. 2012. *1° Magic Tricks Revealed Match One More Time*. Magiamagia.

Quechua. 2011. *2" seconds Air installation and folding*.