English *of* in L1 and L2 speakers' read and spontaneous speech

Helena Spilková and Wim A. van Dommelen
Department of Language and Communication Studies, NTNU

Abstract

This paper compares realizations of the English function word of *in read and spontaneous speech produced by Czech and Norwegian speakers with native productions. Acoustic analysis involved word and segment durations, voicing, formant and band energy measurements. Non-natives appeared to produce longer durations than natives but read productions were longer than spontaneous tokens for natives and non-natives alike. Relative segment durations, however, varied between speaker groups. F1-F0 values in* of *did not differ between groups. Higher F3-F2 values indicated less fronted vowel quality for the non-natives than for the natives. In this case there was no effect of speaking style. Amount of fricative voicing and friction intensity varied systematically with absence or presence of voicing in the following segment.*

Introduction

It is a well-known fact that spoken language is characterized by reduction processes making actual word tokens differ from their canonical forms. Reduction phenomena have been studied in e.g. Dutch (van Son and Pols 1999; Keune et al. 2005), English (Bell et al. 2003; Baker and Bradlow 2009), Greek (Dauer 1980), Japanese (Nakamura et al. 2008), Russian (Bolotova 2003) and Spanish (Harmegnies and Poch-Olivé 1992). One of the most important factors in reduction of word forms is speaking style. Usually, investigations compare more formal presentations (often read speech) with less formal speech (often called spontaneous speech). Although *spontaneous speech* might seem to be appropriate as a cover term, it should be kept in mind that in a number of studies speaking strategies have been shown to be individually different. For example, different approaches to speaking styles were found for the two Dutch speakers in Laan (1997), five speakers of Greek in Dauer (1980), five English subjects in Lavoie (2002), and seven Dutch, six Finnish and five Russian speakers in de Silva et al. (2003).

For second language (L2) speakers it may be a challenge to produce appropriately reduced forms. During the initial period of second language instruction normally attention will be paid to the L2 sound system, in particular to the sounds that are different from the learner's native language (L1). Naturally, much emphasis will be put on the acquisition of specific sounds as they appear in canonical word forms. In a more advanced stage, efforts may be devoted to mastering language-specific reduction rules. Depending on factors like frequency and predictability of a word, its position in an utterance, speaking style, etc., native speakers' realizations may show varying degrees of reduction (cf. Lavoie 2002). Previous investigations have produced evidence of L2 learners showing phrase-level effects that differ from what is found for L1 speakers. Native speakers of Spanish in Lowenstein Mairs (1989) produced incorrect patterns of stress assignment in English. Further, Wenk (1985) found improper reduction of vowel quality in francophone English. In a study by Bond and Fokes (1985), native speakers of Thai, Malaysian and Japanese were shown to have insufficient awareness of typical English patterns of word compression due to addition of syllable suffixes. In Gut (2007) English learners of German demonstrated insufficient degrees of vowel reduction. The same was true for German learners of English.

The goal of the present study was to investigate reduction phenomena in the English function word *of* occurring in (quasi-) spontaneous vs. read speech produced by native as well as non-native speakers. The spontaneous material was collected from different types of task-elicited dialogues recorded in studios (see Method section). Non-native speakers came from two typologically different languages,

Norwegian and Czech (cf. Swan and Smith 2001). It was hypothesized that in general non-native speakers of English would show less reduced productions than natives. Further, native speakers were expected to exhibit stronger reduction effects in spontaneous vs. read speech than L2 users. Finally, it was postulated that the larger typological distance between English and Czech would cause less native-like productions for Czech than for Norwegian speakers.

Method

Speech material

The material used in this study was obtained from several sources. The read material is represented by recordings of non-professional speakers reading transcripts of BBC news texts. Part of the BBC news recordings was recorded in Trondheim and part was provided by the Institute of Phonetics, Charles University in Prague. The spontaneous material consists of spontaneous dialogues in English, elicited using a picture replication task (part of the Kachna corpus; Spilková et al. 2010) and dialogues elicited using a Map Task (White et al. 2010). All the recordings were made in studio environments with a sampling rate of 32 kHz or higher and 16-bit quantization, using a separate channel for each speaker (in dialogues).

The lexical item chosen for analysis is the English preposition *of*. For both types of material, we aimed to select realizations of this word fluently and naturally integrated in surrounding speech, therefore we excluded all cases where a pause, hesitation or another type of disfluency was present in close proximity of the observed word. Attention was also paid to the context and syntactic status of the observed word, where we avoided, e.g., clause-final use of prepositions (so-called stranded preposition) and strongly lexicalized phrases where a disproportional reduction could be expected. Five tokens per speaker and speaking style were selected (incidentally less than five for a few speakers with a limited number of suitable items).

Speakers

The groups of subjects consisted of ten Norwegian speakers (4 females, 6 males), ten Czech speakers (5 females, 5 males) and two native British English speakers (females) that were recorded in both speaking styles (reading and replication task dialogue). In addition, BBC news recordings of three British speakers (males) and recordings of Map Task dialogue of six British speakers (2 females, 4 males) were used. The age of the speakers ranged from 19 to 45 years, and most of the speakers were university students. The speaker pairs in the dialogues were in most cases formed by either classmates or colleagues.

In Norway, the well-established system of English instruction and high exposure to English language (e.g. most movies in English are not dubbed) result in an overall high competence in English in young population. The speakers were therefore selected from university students which guaranteed sufficient proficiency. In Czech Republic, however, such a proficiency standard cannot be generally expected and we had to select speakers from more carefully chosen groups, namely university students of English, and employees in a company using English as the official work language. The dialects of the native English speakers mostly belonged to the Southern English dialect group; one speaker spoke a Northern English dialect.

Acoustic analysis

The selected items were segmented using Praat (Boersma and Weenink 2009). Segment durations were obtained for the vowel, the fricative and the portion with phonetic voicing in the fricative. Formant values in Bark were measured as means of values obtained from the whole duration of the vowel in the observed item. To be able to eliminate relatively frequent errors in automatic formant tracking, an additional semi-automatic method was used to detect any abrupt jumps between nearby formant measurements. The resulting formant values were used to calculate F1-F0 and F3-F2 values in Bark that correspond to vowel height and backness while reducing anatomical variation (corresponding to, e.g., gender; cf. Syrdal and Gopal 1986; Adank et al. 2004). The value of F0 necessary for this transformation was measured in the centre of the vowel interval, avoiding the portions with a creaky voice quality where possible. Furthermore, band energies (low band: 0 - 5000 Hz, high band: over 5000 Hz) were also measured in the fricative. These were used to calculate *high-frequency band – low-frequency band* differences in dB, corresponding to relative friction intensity (that represents the fortis character of the sound).

Results

Segment durations

In this section the results of measurements of segmental durations will be presented. As can be seen from Figure 1, total word durations were longer in read than in spontaneous speech (pooled across the three groups of speakers 123ms vs. 105ms). In addition, word durations were longer for both Czech (133ms) and Norwegian speakers (109ms) compared to natives (91ms). An analysis of variance with speaking style and language as factors revealed that whereas the effects of speaking style as well as language were significant ($F(1, 255)= 6.610$; $p= 0.011$ and $F(2, 255)= 12.601$; $p< 0.001$), the speaking style x language interaction did not reach significance ($F(2, 255)= 1.057$; $p= 0.349$). Bonferroni-adjusted paired comparisons showed that only Czech speakers had significantly longer word durations than natives. Czech word durations were also reliably longer than those produced by Norwegians.

Further, we investigated whether V/C ratios differed across speaking styles and language groups. The data showed that vowel duration in the word *of* in read speech amounted to 50% of the total word duration vs. 55% in spontaneous speech. For the English speakers the corresponding percentages were 47% vs. 55%. While for Czech speakers similar ratios were observed (43% vs. 55%), Norwegian speakers showed an opposite pattern (59% vs. 56%). According to an analysis of variance the effects of the factors speaking style ($F(1, 255)= 6.253$; $p= 0.013$), language ($F(2, 255)= 6.585$; $p= 0.002$) as well as their interaction ($F(2, 255)= 5.396$; $p= 0.005$) were significant.

Closer inspection of the data revealed that speaking style also affected the degree to which the fricative was filled with voicing (69% in read speech vs. 84% in spontaneous speech). For the English and, especially, the Czech speakers this result can be explained by relatively long fricative durations in read speech (read vs. spontaneous: Czech 84ms vs. 56ms; English 58ms vs. 40ms). Norwegian speakers had similar fricative durations for these two conditions (both 48ms). Pooled across the two speaking styles, Czech and English subjects had comparable amounts of voicing (73% and 70%, respectively), while Norwegian speakers made the fricative more voiced (85%). Statistical analysis revealed that the effect of both speaking style and language on the amount of voicing in the fricative was significant ($F(1, 242)= 12.435$; $p= 0.001$ and $F(2, 242= 5.082$; $p= 0.007$) with a significant speaking style x language interaction ($F(2, 242)= 3.875$; $p= 0.022$). Phonetic classification of immediately neighboring segments as voiceless/voiced revealed that the amount of voicing in the fricative correlated with the voicing status of the following segment. This issue will be dealt with in the section Context effects below.

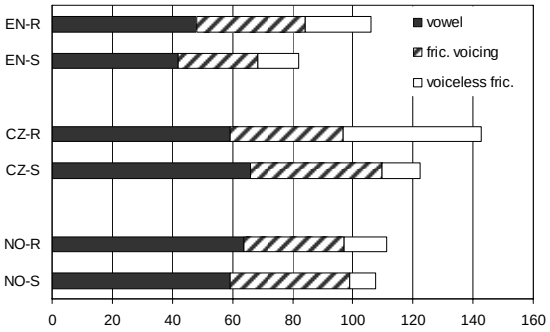

Figure 1. Segment durations in ms in of *spoken by English (EN), Czech (CZ) and Norwegian (NO) speakers. R= read, S= spontaneous; fric. voicing= voiced part, voiceless fric.= voiceless part of fricative.*

Vowel quality

In the following we will investigate whether the quality of the vowel in *of* differed systematically due to speaking style and between natives and non-natives. F1-F0 and F3-F2 values for different conditions are presented in Figures 2 and 3. For F1-F0 an analysis of variance with the factors speaking style and language showed no significant differences at all (for both factors: $F< 1$). Only Norwegian subjects' F1-F0 values were larger for read than spontaneous speech (3.3 Bark vs. 2.9 Bark; $t(92)= 2.168$; $p= 0.033$).

In contrast, for the F3-F2 measure a significant effect of language was found ($F(2, 252)= 24.416$; $p< 0.001$). Bonferroni-adjusted paired comparisons showed that across speaking styles F3-F2 values for Czech (4.6 Bark) as well as Norwegian (4.5 Bark) were different from English (3.6 Bark). This indicates more peripheral vowel qualities for the non-native speakers. Both speaking style (read vs. spontaneous: 4.4 Bark vs. 4.3 Bark) and its interaction with the factor language, however,

did not reach statistical significance (F< 1 and F(2, 252)= 1.668; p= 0.191, respectively).

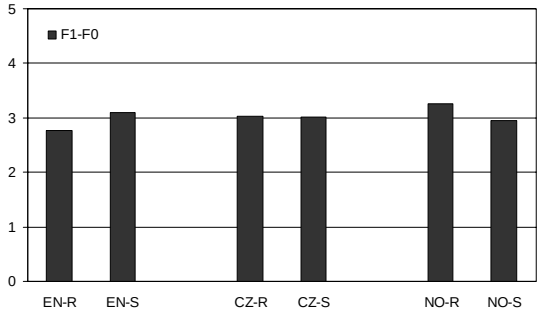

Figure 2. F1-F0 distances in Bark in of. *Legend: see Figure 1.*

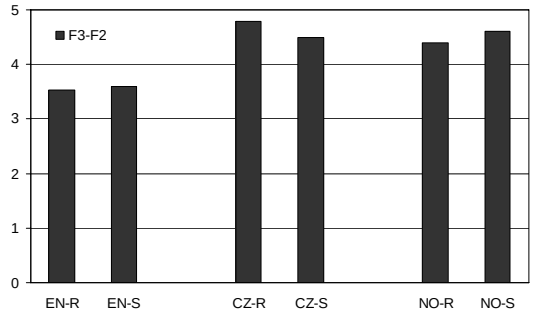

Figure 3. F3-F2 distances in Bark in of. *Legend: see Figure 1.*

Context effects

Fricative voicing

It was speculated that apart from the factors speaking style and speakers' language background, phonetic context may affect realization of the fricative in *of.* To study the effect of context, the immediately neighbouring segments were classified as (phonetically) voiceless or voiced. First we will look into contextual effects on segment durations. As was shown by an analysis of variance with factors left context and right context (in addition to speaking style and language), absence vs. presence of voicing in neither left nor right context did affect total word duration. Also, no significant effect on relative vowel duration was found. In contrast, absence vs. presence of voicing in the segment following *of* had a strong influence on the amount of voicing in the fricative. Across speaking styles and language groups, voicing percentages amounted to 87% before a voiced segment vs. 54% before a voiceless segment (see Figure 4). According to an analysis of variance with factors speaking style, language, left context and right context this effect was highly significant (F(1, 224)= 56.062; p< 0.001). Left context had no significant effect at all (F< 1). The only significant interaction was found between language and speaking style (F(2, 224)= 5.316; p= 0.006). Presumably due to small number of observations in some categories, the main effects of speaking style and language did not reach statistical significance any longer (cf. section Segment durations above).

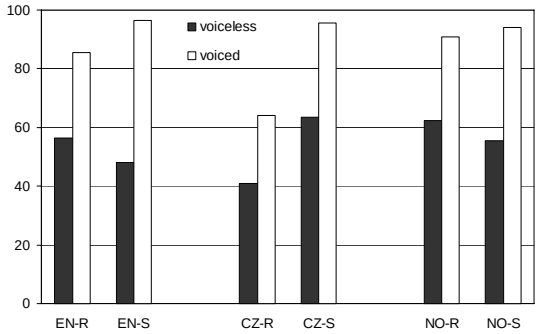

Figure 4. Amount of voicing in the fricative in of *(in %) followed by a voiceless/voiced segment. Legend: see Figure 1.*

Fricative intensity

One of the acoustic correlates of the status of a fricative as fortis vs. lenis is the intensity of the friction. Through a similar analysis as for fricative voicing, in this section we investigate whether the absence vs. presence of voicing in the neighbouring segments also affected friction intensity. The measure used was the difference in spectral energy above-below 5000Hz in the fricative (see Method section, Acoustic analysis).

As can be seen from the results presented in Figure 5, intensity of the friction varies with voicing status of the following segment. When followed by a voiceless segment, the amount of spectral energy above 5000Hz is considerably larger than preceding a voiced segment. An analysis of variance with factors speaking style, language, left context and right context showed that the latter effect was highly significant (F(1, 224)= 52.742; p< 0.001). The same was true for the main effects speaking style and language (F(1, 224)= 7.856; p= 0.006 and F(2, 224)= 3.887; p= 0.022) but not for left context (F(1, 224)=1.04; p= 0.295. The only significant

interaction was between speaking style and language ($F(2, 224)= 5.975$; $p= 0.003$).

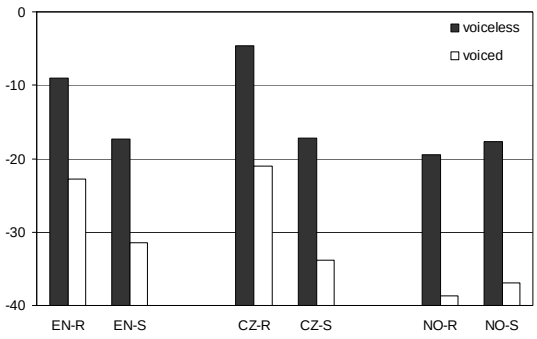

Figure 5. Difference in spectral energy above-below 5kHz in the fricative in of *(in dB) followed by a voiceless/voiced segment. Legend: see Figure 1.*

Discussion

The combined results of the present acoustic measurements on read and spontaneous *of* tokens produced by L2 speakers revealed partly consistent tendencies on the one hand and partly diverging reduction patterns on the other. In the temporal domain there was a general tendency to longer word durations for the non-native speakers than for the natives (though only significant for the Czech). This result is in line with Gut's (2007) observations of longer syllable durations in non-native English spoken by Germans vs. native English and *vice versa* in non-native German produced by English speakers. Also, White and Mattys (2007) found speech rate to be slower for non-natives compared to natives in Spanish sentences read by English speakers as well as English sentences read by Spanish speakers. In the study by Mackay and Flege (2004) early and late Italian–English bilinguals repeated matched English and Italian sentences following an aural model. The late bilinguals produced longer English than Italian sentences. This result was interpreted as due to speakers' need for resources to suppress their Italian subsystem. A similar interpretation might be valid for the present results.

Also in the spectral domain a general L1-L2 effect was observed. As evidenced by the F3-F2 measure, the Czech as well as Norwegian subjects produced both read and spontaneous *of* with a less reduced vowel than the natives. This result is in congruence with Flege, Bohn and Jang (1997) who observed inappropriate spectral contrasts in English vowel pairs produced by German, Mandarin, Spanish and Korean speakers. In Wenk (1985) French L2 users of English showed insufficient reduction of vowels in pre-tonic syllables. Lowenstein Mairs (1989) reports inappropriate stress assignment by Spanish speakers of English, which presumably caused insufficient vowel reduction (neither her nor Wenk's study involved acoustic measurements).

At odds with the general differences between L1 and L2 performance was the absence of significant interactions for language and speaking style. Word durations were shorter in spontaneous than in read speech for natives and non-natives alike. Further, no vowel quality reduction due to speaking style was found for L1 as well as L2 speakers. For the English speakers this might be due to high degree of vowel quality reduction present already in their read tokens.

Also the effects of context on *of* realizations were similar for the native and non-native speakers. For all three groups, absence vs. presence of phonetic voicing in the segment following *of* affected voicing during the fricative and also its intensity. It can be speculated that we are dealing with assimilation processes that are possibly similar for the three languages (cf. Hall 2003 for Czech; Jansen 2007 for English; not many studies seem to exist for Norwegian; cf. Kristoffersen 2000). More research on assimilation across word boundaries is needed to answer this question.

Further, internal syllable structure (i.e., V/C ratio) differed for natives vs. Norwegian but not Czech speakers. Only the Norwegians had a larger percentage of voicing in the fricative which was explained by their relatively short fricative durations.

The lack of consistent results might at least partly be explained by individually different reduction strategies. It might be speculated that in the present study idiosyncratic behaviour to some degree outweighed L2-specific influences.

Acknowledgements

The authors were supported by the EU Marie Curie Research Training Network Sound to Sense. We would like to thank Radek Skarnitzl for providing the read Czech material and to Laurence White for making available the recordings of native English speakers' Map Task.

References

Adank P, Smits R, van Hout R (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America,* 116: 3099-3107.

Baker R E, Bradlow A R (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech,* 52: 391-413.

Bell A, Daniel J, Fosler-Lussier E, Girand C, Gregory M, Gildea D (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America,* 113: 1001-1024.

Boersma P, Weenink D (2009). Praat: Doing phonetics by computer (Version 5.1.06) [Computer program]. Retrieved May 11, 2009, from http://www.praat.org/.

Bolotova O (2003). On some acoustic features of spontaneous speech and reading in Russian (quantitative and qualitative comparison methods). *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003 (1): 913-916.

Bond Z S, Fokes J (1985). Non-native patterns of English syllable timing. *Journal of Phonetics*, 13: 407-420.

Dauer R M (1980). The reduction of unstressed high vowel in modern Greek. *Journal of the International Phonetic Association,* 10(1): 17-27.

de Silva V, Iivonen A, Bondarko L V, Pols L C W (2003). Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003 (3): 2977-2980.

Flege J E, Bohn O-S, Jang S (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25: 437-470.

Gut U (2007). Learner corpora in second language prosody research. In: Trouvain J, Gut U, eds, *Non-native prosody: Phonetic description and teaching practice.* Berlin/New York: Mouton de Gruyter, 145-167.

Hall D C (2003). Laryngeal feature specifications in West Slavic languages. *Toronto Working Papers in Linguistics,* 20: 93-114.

Harmegnies B, Poch-Olivé D (1992). A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish. *Speech Communication,* 11: 429-437.

Jansen W (2007). Phonological 'voicing', phonetic voicing, and assimilation in English. *Language Sciences,* 29: 270-293.

Keune K, Ernestus M, Van Hout R, Baayen, R H (2005). Variation in Dutch: From written MOGELIJK to spoken MOK. *Corpus Linguistics and Linguistic Theory*, 1: 183-223.

Kristoffersen G (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.

Laan G P M (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication,* 22: 43-65.

Lavoie L (2002). Some influences on the realization of for and four in American English. *Journal of the International Phonetic Association*, 32(2): 175-202.

Lowenstein Mairs J (1989). Stress assignment in interlanguage phonology: an analysis of the stress system of Spanish speakers learning English. In: Gass S M, Schachter J, eds, *Linguistic perspectives on second language acquisition.* Cambridge: Cambridge university Press, 260-283.

Mackay I R A, Flege J E (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, 25: 373–396.

Nakamura M, Iwano K, Furui S (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language,* 22: 171-184.

Spilková H, Brenner D, Öttl A, Vondřička P, van Dommelen W A, Ernestus M (2010). The Kachna L1/L2 Picture Replication Corpus. *International Conference on Language Resources and Evaluation*, 2010, Valletta. (to appear)

Syrdal A K, Gopal H S (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79: 1086-1100.

Swan M, Smith B, eds, (2001). *Learner English: a teacher's guide to interference and other problems.* (2nd ed.) Cambridge: Cambridge University Press.

van Son R J J H, Pols L C W (1999). Acoustic description of consonant reduction. *Speech Communication,* 28: 125-140.

Wenk B J (1985). Speech rhythms in second language acquisition. *Language and Speech,* 28: 157-175.

White L, Mattys S L (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35: 501–522.

White L, Wiget L, Rauch O, Mattys S L (2010). Segmentation cues in spontaneous and read speech. *Proceedings of the 5th Conference on Speech Prosody*, 2010, Chicago. (to appear)