

Differences in codeswitching patterns between immigrant and heritage speakers of US American Danish

Karoline Kühl & Jan Heegård Petersen*

Abstract. This paper addresses intrasentential code-switching in US American Danish spoken by 1st generation immigrants from Denmark, who migrated to the US around 1900 as part as the transatlantic European mass emigration and their descendants (2nd and 3rd generation). The analysis is based on the *Corpus of American Danish*, specifically a dataset of 173 speakers producing 46 hours of speech. In this dataset, we observe significantly different patterns of intrasentential code-switching in the speech of the immigrant speakers (1st generation) and US-born heritage speakers (2nd and 3rd generation): The code-switching patterns of the heritage speakers show a preference for English lexemes that are integrated morphologically into Danish or which are part of Danish-English bilingual compounds. In contrast, the immigrant speakers prefer non-integrated English words for code-switching. This result taken per se shows that code-switching patterns show variation across generations just as other linguistic variables. Taking the result further, we have connected it to a previous study of representations of linguistic proficiency in immigrant Danish in the Americas. By this, we are able to show that the morphologically or lexically integrated code-switching of the heritage speakers correlate with features representing fluent speech, while the non-integrated code-switching of the immigrant speakers rather seems to correlate with features showing a low activation of Danish in a situation of language shift. Thus, the heritage speakers seem to have developed a way of speaking US American Danish where English word stems are an integrated part of speaking fluently and lexically varied.

Keywords. US American Danish; heritage language; minority language; immigration; language contact; codeswitching

1. Introduction. The Danes who came to the United States of America as part of the European mass emigration around 1900 and their descendants have mostly lived a life surrounded by English as the majority language. Hence, it comes as no surprise that US American Danish is characterized by the occurrence of English words, word stems and multi-word phrases. The reasons for codeswitching are notoriously numerous: Switching to another language for a word, a word stem or a phrase may come about as the result of lexical retrieval difficulties, it may represent unmarked language production in a bilingual context where there is no need to stick to monolingual language production norms, or the switch might aim at stylization for pragmatic reasons (see, e.g., Auer & Eastman 2010). Working with speech produced by descendants of immigrant speakers, we also have to keep in mind that the parent generation may already have incorporated words and phrases from the majority language into their variety, implying that the next generations may never have heard the equivalent in the homeland language (Riehl 2015:277) or that the switch may be a cultural loan that is strongly associated with the recipient country, e.g., new social activities, institutions, fauna and flora, and therefore difficult to translate without losing meaning (Matras 2009:150).

* We would like to thank the peer-reviewers and the editors for valuable comments to a previous version of the paper. Authors: Karoline Kühl, University of Flensburg (karoline.kuehl@uni-flensburg.de) & Jan Heegård Petersen, University of Copenhagen (janhp@hum.ku.dk).

The purpose of this study is to identify patterns of different kinds of codeswitching in two different groups of speakers of US American Danish: 1st generation immigrant speakers who settled in the United States on average in early adulthood, and 2nd and 3rd generation heritage speakers who were born and raised in the United States. By ‘different kinds of codeswitching’ we do not, however, mean different reasons for codeswitching. Rather, this study is a structural study of codeswitching, focusing on different degrees of adaptation of the English elements into Danish (cf. below) related to the two different speaker groups mentioned above. This paper combines an empirical analysis of codeswitching patterns based on the *Corpus of American Danish* with a partial result from a previous publication on US American and Argentine Danish (Heegård Petersen et al. 2018).

Terminology with regard to denoting the use of two (or more) languages within one utterance is also quite varied, e.g., codeswitching, codemixing (Auer & Eastman 2010), insertion or alternation (Muysken 2000). In this study, we will use ‘codeswitching’ as a merely descriptive term, denoting the occurrence of English words or word stems in otherwise Danish utterances. We only address intrasententially occurring codeswitching. This matches Muysken’s category *insertion* which denotes lexical elements embedded in another language (Muysken 2000, ch. 3). In this study, we distinguish between the following kinds of codeswitching/insertions in the following examples, English elements are marked by bold type:

a. Insertion of English words not adapted morphologically to Danish as shown in (1) and (2).

- (1) vi skrev en gang imellem, sendte et **picture** over til Danmark
 ‘We wrote once in a while, sent a picture to Denmark.’
- (2) de kunne snakke fem **different** sprog
 ‘They could speak five different languages.’

b. English word stems morphologically adapted to Danish as in (3).

- (3) en lille forretning som vi **rent**-ede
 a small store that we rent-PST
 ‘A small store that we rented.’

In (3), the English verb stem *rent* is combined with the regular Danish preterite ending *-ede*. In this category, we mostly find English noun stems or verb stems, but occasionally also adverbs and adjectives.

c. Bilingual composita containing two nouns as in (4)–(6). The bilingual compounds may also be morphologically integrated into a Danish inflection pattern as shown in (4) and (5).

- (4) **majsharvest**-en
 corn_harvest-DEF
 ‘the corn harvest’
- (5) **milkvogn**-e
 milkvan-PL
 ‘dairy vans’
- (6) **skomager**shop
 cobbler_shop
 ‘cobbler’s shop’

In order to distinguish categories of codeswitching, we have termed cases like (1)–(2) that show no integration into Danish, ‘English words’, whereas cases like (3)–(6) are called ‘integrated codeswitching’. In this particular study, these differences regarding the adaptation to Danish are important as they are linked with intergenerational differences. This will be shown below.

The question of when and how ad hoc codeswitching turns into established loan words (often termed ‘borrowing’) is highly interesting and, of course, relevant for this study. Poplack has offered a number of criteria addressing both formal and functional characteristics of loan words as opposed to ad hoc-codeswitching or borrowing, which might be summarized as follows (Poplack 2017:122–140):

- i. loan words assume the morphological, syntactic, and sometimes the phonological identity of the recipient language
- ii. they occur recurrently in the speech of the individual (1 type to many tokens) and are also used widely across the community
- iii. established loanwords are available to monolingual speakers of the recipient language, who access them normally along with the remainder of the recipient-language lexicon.

As clear as these criteria may seem, to operationalize them on spoken, non-codified moribund or already extinct heritage varieties is challenging. The formal criteria (i.) regarding the degree of integration into the recipient language may be operationalized as in this study by assessing the patterns of morphological and lexical integration of the codeswitching (English) into the receiving language (Danish). Regarding the degree of establishment across individuals and the community (ii.), this might be done for the individual speakers contained in the dataset. We cannot, however, establish the use of potential loanwords in the community because the recordings contained in the dataset (the US American part of the *Corpus of American Danish*, cf. Section 2) only partly seem to represent social networks.¹ Testing the availability of potential English loan words to monolingual speakers of (contemporary) Danish (iii.) is impossible due to the historical nature of the data. Another question not easily answered would be which variety would count as the recipient language: the homeland variety or US American Danish? Speakers of US American Danish would by definition not be monolingual, with at least receptive competence in English.

In a previous study, we observed the number of recurring codeswitches (or loan words) in almost the whole US American Danish dataset (regardless of social networks) to be quite low: Only 3.5% of all English words or word stems occur in a relation of 1 type and many tokens. Not unexpectedly, they predominantly cover cultural loans such as ‘dollar’, ‘acre’ and ‘miles’ that hardly can be avoided (Heegård Petersen, Kühl & Hansen 2020:100–101). Based on this knowledge, we will in this study not go into details with regard to whether the occurring English words have become loan words or if they still should be considered ad hoc switches. The criterion applied to determine the language association of a word or a word stem is simply whether the word occurs in *Den Danske Ordbog* (‘The Danish Dictionary’), which covers the period 1954 till present, the *Ordbog over det danske sprog* (‘Dictionary of the Danish Language’), which covers the period 1780–1954, or the Danish dialect dictionaries.

This paper is structured as follows: In Section 2, we present the *Corpus of American Danish* and the data subset that this study is based on. Section 3 presents the empirical analysis of codeswitching patterns related to the speaker groups, while Section 4 provides a short review of relevant results from a previous study of US American Danish. In Section 5, we connect and discuss the results, and Section 6 is a short conclusion.

2. Data. The *Corpus of American Danish*. The analysis is based on the *Corpus of American Danish*, which today is hosted by the LANCHART corpus at the University of Copenhagen (Kühl, Heegård Petersen & Hansen 2020). The corpus consists of recordings of Danish speech

¹ Only few of the recordings contain remarks by the interviewees establishing that they knew other informants. Original field notes are lost (if they ever existed).

produced by immigrant and heritage speakers in Argentina and North America (the United States and Canada) and as such represents multilingual spoken language.

The recordings of US American Danish that are included in this study were obtained between 1963 and 1982 in various places in the United States by the late Danish linguists Iver Kjær and Mogens Baumann Larsen. Originally, Kjær and Baumann Larsen’s aim was to identify fossilized dialectal features and features of older Danish, but quite soon, they became interested in US American Danish in its own right (Baumann Larsen & Kjær 1978). They conducted sociolinguistic interviews centered around a number of recurring topics such as the situation in Denmark before the immigration, the immigration process itself, the Great Depression and autobiographical narratives of life in the United States. No conscious decisions around language choice by Kjær and Baumann Larsen have been documented, but our impression from the interviews is that the informants knew the interviewers to be competent in English and all agreed that Danish should be the language for the interview. This does not prevent either codeswitching nor occasional slips into English. Ultimately, the conversations always return to Danish.

The data have been transcribed, annotated and connected to the sound files. As of today, the *Corpus of American Danish*, including data from Argentina, Canada and the USA, contains approx. 1.3 million tokens (word forms, but also phenomena of spoken language such as laughter, re-starts, hesitation phenomena etc.) produced by ca. 300 speakers. Including the interviewer’s speech, the corpus amounts to approx. 1.7 million tokens; Kühl et al. (2020). The dataset used for this particular analysis is restricted to data from US American Danish speakers. Table 1 below provides an overview.

Speakers	173 (80 men, 93 women)
Birth year	1870–1939
Age when recorded	40–97, median age 87
Birthplace	Denmark (immigrant speakers): 119 North America (heritage speakers): 54
Age of immigration (Danish-born speakers)	0–44, median age 21 ²
Hours of recorded speech	46

Table 1. Dataset of US American Danish extracted from the *Corpus of American Danish*

Table 1 shows that the dataset of this study consists of the recordings of in total 173 US American Danes, amounting to 46 hours of speech. 119 of the speakers were born in Denmark; these we term ‘immigrant speakers’. Their age by the time of immigration ranges from 0-44 years, but as can be seen from the median age (21 years), most of them emigrated as young adults (see also footnote 2). This implies that they all had acquired Danish as their only language prior to emigration and had acquired literacy in Danish (although those who worked as farmhands in Denmark would have had just a few years of schooling).

The *Corpus of American Danish* has been coded throughout for a number of linguistic variables, among others the language of words, word-internal codeswitching, empty and filled pauses, repetition, self-interruption and syntactic information, specifically Danish main clauses and subclauses along with the degree of syntactic subordination. For the purpose of ascribing

² Only 14 immigrants are below the age of 14 when it would have been acceptable for children from working class families to leave school.

a base language to a sentence, we have defined a sentence as Danish if the grammatical subject and the finite verb is in Danish as in (7) below.

- (7) Jeg er en **American citizen**
 ‘I am an American citizen.’

If the grammatical subject and the finite verb is in English (as in (8) below) this sentence would count as an English sentence (cf. Kühn et al. 2020 for an overview of the coding).

- (8) **I am an amerikansk statsborger.**
 ‘I am an American citizen.’

Ascribing a base language to an utterance by determining the language of the grammatical subject and finite verb should not be confused with the rather more sophisticated approaches towards determining the matrix language and defining the ensuing possibilities of codeswitching presented by, e.g., Myers-Scotton (1993), Jake & Myers-Scotton (2009), and Muysken (2000). The principle used here is an operational approach towards creating a meaningful unit for the analyses of codeswitching, not a categorization based on a theoretical framework. Based on this coding, we extracted the number of Danish words, English words and occurrences of integrated codeswitching that occur in these Danish sentences and from that, we analyzed code-switching patterns between the two groups of speakers.

3. Analysis: Codeswitching patterns in immigrant and heritage speakers of US America Danish. Table 2 shows the quantitative distribution of Danish words, English words and integrated codeswitching containing an English word stem for all speakers (1st column) and specifically for the Danish-born immigrant speakers (2nd column) and US-born heritage speakers (3rd column).

	All speakers (n=173)	Immigrant speakers (n=119)	Heritage speakers (n=54)
Danish words	267,859 (100%)	173,682 (65%)	94,177 (35%)
English words	31,458 (100%)	25,656 (81%)	5,802 (19%)
Integrated codeswitching	578 (100%)	341.0 (59%)	237 (41%)

Table 2. Occurrence of Danish words, English words and integrated codeswitching in Danish sentences produced by immigrant speakers and heritage speakers

Based on the observed numbers, we have used a chi square-test to determine whether the differences in the occurrence of Danish and English words/word stems in the speech of immigrant and heritage speakers, respectively, are statistically significant. More specifically, we have looked for statistical significance in the following combinations (always relating to immigrant and heritage speakers as two different speaker groups):

- Danish words vs. English words: $p < .00001$ ($\chi^2 = 3535.9831$; df 1, N = 299,317),
- Danish words vs. integrated codeswitching: $p = .003289$ ($\chi^2 = 8.6397$; df 1, N = 268,437),
- Danish words and English words vs. integrated codeswitching: $p = .000109$ ($\chi^2 = 14.9799$; df 1, N = 299,895),
- Danish words and integrated codeswitching vs. English words: $p < .00001$ ($\chi^2 = 3541.3368$; df 1, N = 299,895).

The statistical significance of all these combinations implies that the quantitative differences in the use of codeswitching (English words and integrated codeswitching, respectively) in the two different speaker groups (heritage vs. immigrant speakers) are not caused simply by chance.

The table and the tests for significance reveal interesting differences: The immigrant speakers produce more English words, but less integrated codeswitching than the heritage speakers. In other words, the number of integrated codeswitching occurring in the heritage speaker group is high compared to the immigrant speakers. This means that the US-born heritage speakers of US American Danish are more likely to produce morphologically integrated codeswitching (such as English verb stems combined with Danish tense marking or English noun stems combined with the Danish enclitic marker of definiteness) and bilingual composita in their Danish speech than the immigrant speakers are. The immigrant speakers are more likely to produce codeswitching in form of non-morphologically adapted English words.

We have controlled for the fact that either of the groups might produce less Danish which in turn would mean less contexts where English codeswitching might occur. However, the number of Danish words produced on average by immigrant and heritage speakers are comparable: On average, the immigrant speakers produce approx. 1460 Danish words, while the heritage speakers produce approx. 1740 words (cf. Table 2).

In turn, this means that the intergenerational differences in the codeswitching patterns in US American Danish are, in fact, not an effect of methodological decisions. Thus, we observe different speech patterns in the immigrant and heritage speakers: The immigrant speakers prefer English words that are not adapted to Danish for codeswitching (cf. the examples (1) and (2) above), the heritage speakers prefer to adapt English lexemes either morphologically or lexically by constructing bilingual compounds (cf. examples (3)-(6) above). These differences resonate with a result from a previous study on US American Danish which we will account for briefly in the following section.

4. Review of results from Heegård Petersen et al. (2018). A partial result from a previous publication on the immigrant and heritage varieties of Danish in the Americas is particularly interesting regarding the study of codeswitching patterns as related to generation. Heegård et al. (2018) is a study of linguistic proficiency which has been operationalized by analyzing the clustering of 13 linguistic variables through Factor Analysis. The study does not access the degree of linguistic proficiency but rather the way these variables cluster and thereby create or represent patterns of linguistic proficiency. Interestingly, the study showed a large degree of consistency between the US American immigrant speakers and the heritage speakers.

For both immigrant and heritage speakers, we found the following: Filled pauses, empty pauses, self-interruption, lengthening, speech rate (Danish words pr. second) and run length (i.e., the length of an utterance in terms of number of words excluding pauses, self-interruptions etc.) cluster into one factor (i.e., a grouping of variables), but with speech rate and run length loading negatively. This tells us that if a speaker has many pauses, self-interruptions and lengthening, his/her speech rate will be slow, and their run length comparatively slow, which is an intuitively plausible result. Hence, this factor represents hesitation and disfluency which, of course, is an aspect of linguistic proficiency. The variables English words and integrated codeswitching cluster with ratio of subclauses to main clause and Danish words into another factor, with the latter two variables loading negatively. In other words, if speakers produce many English words and integrated codeswitching, they will have fewer Danish words and a lower degree of syntactic subordination. This factor seems to represent an ability to keep languages apart or maybe execute cognitive control over speech. The remaining variables, i.e., type-token ratio, word length and repetition, cluster into a third factor, with repetition loading negatively. This tells us that if speakers have a high type-token ratio, they will also produce longer sentences and fewer repetitions. This factor seems to represent proficiency in terms of a rich lexicon.

The only difference between heritage and immigrant speakers in this study is the way in which the variable ‘integrated codeswitching’ clusters with other variables: For the immigrant

speakers, it clusters with many English words, few Danish words, and a low degree of syntactic subordination. These variables taken together seem to represent a low degree of proficiency in Danish. For the heritage speakers, though, integrated codeswitching correlates with a high type-token ratio, longer sentences and fewer repetitions. This, in turn, represents a well-developed lexicon and fluent speech (cf. Heegård Petersen et al. 2018:16–19). This partial result from Heegård Petersen et al. (2018) will be connected to the results from the analysis presented here in the following Section 5.

5. Discussion. The analysis in Section 3 has shown that the speech production of Danish-born immigrant speakers and US-born heritage speakers of US American Danish differs significantly: The immigrant speakers produce more English words in Danish sentences and less integrated codeswitching than the US-born heritage speakers.

This result seems counterintuitive. As pointed out in Section 2, the immigrant speakers represented in the *Corpus of American Danish* emigrated as young adults. They had acquired Danish probably as their only language prior to emigration, implying a monolingual acquisition of Danish through the formative years. In contrast, the heritage speakers of US American Danish have been under considerable influence of English for the whole of their lives, particularly because the US American Danes by and large did not form particularly tight networks among themselves, and language shift seems to have been fairly advanced already in the first US-born generation (Kühl 2015). Hence, language attrition at the individual level and differential acquisition outcomes due to missing input and lack of formal education in Danish (see, e.g., Putnam, Kupisch & Pascual y Cabo 2018:262) most probably influenced the Danish language competence of the heritage speakers. As such, one would have expected the heritage speakers' Danish language competence to be less stable and, accordingly, would have expected them to be more susceptible to produce codeswitching, regardless of the form that the codeswitching takes.

However, the heritage speakers do not produce simply more codeswitching, they show a different pattern. This resonates with the results by Heegård et al (2018) reported above, where integrated codeswitching clusters together with linguistic variables representing linguistic fluency and a rich lexicon for the heritage speakers. For immigrant speakers, integrated codeswitching correlates with English words and lesser degree of syntactic subordination which rather seem to represent a language shift process where speaking Danish has become a laborious attempt rather than fluent speech. These results provide us with a more nuanced picture of the transmission process of immigrant languages: Despite the fact that US American Danish did not survive for much more than two generations (Kühl 2015), the intermediate stages were not simply characterized by an ever-growing amount of English codeswitching in the Danish speech. Rather, the heritage speakers seem to have developed a way of speaking where English word stems are an integrated part of speaking US American Danish fluently and lexically varied.

6. Conclusion. We conclude that it is rewarding to take intergenerational differences into account when analyzing immigrant and heritage languages. This implies once again that it is necessary to establish the baseline for language acquisition of heritage speakers (here, the codeswitching patterns of the immigrant speakers) in order to achieve an adequate means of comparison. However, this study has shown once again that speech patterns (here, codeswitching patterns) are not just passed on as the next generation – here, the heritage speakers – may transform the input that they have received into something different. This implies, ultimately, that codeswitching cannot be judged by itself as, e.g., a manifestation of either high or low heritage language competence in a process of intergenerational language

shift, but that different forms of codeswitching may be or become part of a fluently spoken register or variety. Ultimately, codeswitching should be considered a linguistic variable.

The identification of these qualitative and quantitative patterns in this study and in the study by Heegård et al. (2018) has only been possible by the exploitation of the *Corpus of American Danish*, showing that valuable insights into the development of immigrant and heritage minority languages can be obtained from large-scale corpus linguistic studies.

References

- Auer, Peter & Carol Eastman. 2010. Codeswitching. In Jürgen Jaspers, Jef Verschueren & Jan-Ola Östman (eds.), *Society and language use* (Handbook of pragmatics highlights 7). 84–112. Amsterdam: John Benjamins.
- Baumann Larsen, Mogens & Iver Kjær. 1978. Norms in Danish-American. In John Weinstock (ed.), *The Nordic Languages and Modern Linguistics 3. Proceedings of the Third International Conference of Nordic and General Linguistics*. 192–194. Austin.
- Heegård Petersen, Jan, Karoline Kühl & Gert Foget Hansen. 2020. Kodeskift i de udvandrerdanske varieteter i Nordamerika og Argentina: Inventar, frekvens, funktion. *NyS – Nydanske sprogstudier* 58. 79–112. <https://doi.org/10.7146/nys.v1i58.120484>.
- Heegård Petersen, Jan, Jacob Thøgersen, Gert Foget Hansen & Karoline Kühl. 2018. Linguistic proficiency in immigrant and heritage speakers of Danish in Argentina and North America: A quantitative approach. *Corpus Linguistics and Linguistic Theory* 17(2). <https://doi.org/10.1515/cllt-2017-0088>.
- Kühl, Karoline. 2015. 'Det er easy at tale engelsk også'. Amerikadansk i 1960'erne og 1970'erne. *NyS – Nydanske sprogstudier* 47. 39–64. <https://doi.org/10.7146/nys.v47i47.19875>.
- Kühl, Karoline, Jan Heegård Petersen & Gert Foget Hansen. 2020. The *Corpus of American Danish*: A language resource of spoken immigrant Danish in North and South America. *Language Resources and Evaluation* 54(3). 831–849. <https://doi.org/10.1007/s10579-019-09473-5>.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Muysken, Pieter. 2000. *Bilingual speech. A typology of code-mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Oxford University Press.
- Jake, Janice & Carol Myers-Scotton. 2009. Which language? Participation potential across lexical categories in codeswitching. In Ludmila Isurin, Donald Winford & Kees de Bot (eds.), *Multidisciplinary approaches to code switching* (Studies in bilingualism 41). 207–242. Amsterdam: John Benjamins.
- Poplack, Shana. 2017. *Borrowing: Loanwords in the Speech Community and in the Grammar*. New York: Oxford University Press.
- Putnam, Michael T., Tanja Kupisch & Diego Pascual y Cabo. 2018. Different situations, similar outcomes. Heritage grammars across the lifespan. In David Miller, Fatih Bayram, Jason Rothman & Ludovica Serratrice (eds.), *Bilingual cognition and language. The state of the science across its subfields* (Studies in bilingualism 54). 251–279. Amsterdam: John Benjamins.
- Riehl, Claudia Maria. 2015. Language attrition, language contact and the concept of relic variety: the case of Barossa German. *International journal of the sociology of language* (236). <https://doi.org/10.1515/ijsl-2015-0028>.
- Den Danske Ordbog ('The Danish Dictionary'): ordnet.dk/ddo
Ordbog over det danske sprog ('Dictionary of the Danish language'): ordnet.dk/ods