

Same, same, but *erm sort of* different? Comparing three kinds of fluencemes across Australian, British, Canadian, and New Zealand English

Karola Schmidt^a – Sandra Götz^b – Katja Jäschke^c – Stefan Th. Gries^{ad}

Justus Liebig University Giessen^a / Germany
Philipps University Marburg^b / Germany
University of Siegen^c / Germany
University of California, Santa Barbara^d / United States

Abstract – Although L1-English fluency has been extensively studied from many angles, few contrastive studies examine whether fluency develops similarly or differently across L1-varieties while taking sociolinguistic variation into consideration. This paper aims to close this research gap and examines the use of three core strategies of fluency (or fluencemes), i.e. discourse markers, filled pauses and unfilled pauses, across Australian, British, Canadian, and New Zealand English. These fluencemes were extracted and manually disambiguated from the private conversation sections of the respective components of the *International Corpus of English* (ICE-AUS, ICE-GB, ICE-CAN, and ICE-NZ). The data were normalised per speaker and linked with the sociobiographic metadata of the speakers. Analysis using random forests revealed a consistent fluenceme distribution across the four varieties, with unfilled pauses being the most common, followed by discourse markers, and then filled pauses. This pattern suggests a ‘common fluenceme core’ among L1-English varieties. The influence of sociolinguistic variables —gender, age, education, and occupation— was modest across varieties and exhibited diverse trends. Male speakers tend to use filled pauses more frequently but fewer unfilled pauses compared to female speakers. Increasing age did not significantly affect the frequency of these strategies; however, older speakers tend to use discourse markers less frequently. Both education and occupation showed a slight positive correlation with overall fluency.

Keywords – fluency; filled pauses; unfilled pauses; discourse markers; spoken English; inner-circle varieties of English



1. INTRODUCTION¹

Fluency in English has been widely researched in the past few decades, with previous studies having mainly taken psycholinguistic, cognitive, or sociolinguistic perspectives (e.g. Goldman-Eisler 1961; Albert 1980; Tottie 2015; Crible 2018; Beier *et al.* 2023). It is common knowledge that L1-speakers “violate the idealised rules of spontaneous speech while the discourse itself is still (most generally) perceived as fluent” (Dumont 2018: 63; see also Corley *et al.* 2007). Studies indicate that speakers usually bridge the gap between online processing demands and speaking by using different kinds of fluencemes, that is, “abstract and idealized feature[s] of speech that contribute [...] to the production and perception of fluency, whatever [...] [their] concrete realization might be” (Götz 2013: 8). Fluencemes can thus be used as strategies to overcome necessary planning phases in speech and can be realised, for example, as filled pauses (such as *er, erm*), unfilled pauses (i.e. pauses that are not filled with a non-verbal sound), discourse markers (such as *you know* and *like*, etc.), “smallwords” (Hasselgren 2002) (such as *sort of/sorta, kind of/kinda*), repeats (e.g. *it’s it’s it’s difficult*), incomplete utterances, etc. The necessity to use such planning strategies has been widely recognised in earlier research on fluency, for example by Beeching (2016: 100), who argues that “[p]ragmatic markers are perfectly adapted to the linear online editing which is required in spontaneous speech,” or by O’Connell and Kowal (see 2005: 572) in their work on filled pauses. Despite this general consensus about the frequent presence of fluencemes in L1-English speech, research has not yet revealed whether their frequencies and distributions are similar or different across L1-varieties of English. Accordingly, contrastive studies investigating such potential differences have only been rarely conducted, and, if so, mainly compared only two L1-varieties and/or only one fluenceme (e.g. Miller 2009 on discourse markers in Australian vs. New Zealand English, or Tottie 2011, 2015 on the use of filled pauses as ‘planners’ in British vs. American English). Moreover, regarding potential sociolinguistic variation in the use of fluencemes, previous research has predominantly concentrated on a single variety of English (e.g. Weiss *et al.* 2006; Laserna *et al.* 2014; Fruehwald 2016;

¹ We would like to gratefully acknowledge that this research project has been generously funded by the German Research Foundation (DFG, Reference Numbers GO 1760/4-1 and WO 2224/1-1) as part of a larger project on “Fluency in ENL, ESL and EFL: A contrastive corpus-based study of English as a first, second, and foreign language.” We would also like to thank Christoph Wolk for programming the data coding app and our student assistants Lara Möller, Hannah Vehrs, and Daniel Walker for their invaluable help with the data coding and their patience with long discussions over fluencemes. All remaining errors and infelicities are, however, our responsibility alone.

Scheuringer *et al.* 2017; Sokołowski *et al.* 2020; the studies in Leuckert and Rüdiger 2021). Whether or not there is sociolinguistic variation in fluenceme use across L1-varieties is still largely unknown. Furthermore, while there is a substantial body of research on the use of individual fluencemes, it tends to focus heavily on either discourse markers or filled pauses (see Section 2). Studies exploring fluencemes in combination, however, remain notably scarce. To the best of our knowledge, no contrastive research has yet examined the combination of different fluencemes across English varieties while also addressing sociolinguistic variation.

In this paper, we aim to advance the cumulative understanding of fluency by addressing these research gaps. In doing so, we examine how L1-English speakers from four L1-varieties of English use fluencemes to navigate planning phases, considering a range of social variables. Accordingly, this paper seeks to address the following research questions:

- 1) Are there differences in fluenceme use between different L1-varieties of English?
- 2) Do sociolinguistic variables play a role in predicting the choice of particular fluencemes across varieties?

The remainder of this paper is thus structured as follows. After giving an overview of previous research on fluency and fluencemes in L1-English (Section 2), we present the database used and methods applied (Section 3). In Section 4, we present the findings of our analysis and then round up this paper with a discussion, a conclusion, and an outlook to future research in Section 5.

2. FLUENCY IN L1-VARIETIES OF ENGLISH

Although fluency boasts a relatively long research tradition in linguistics, there are considerable differences in how the concept is defined and operationalised. There seems to be a consensus that fluency is linked to swift and effortless speech production (see Chambers 1997: 535). On a conceptual level, fluency is, however, more frequently linked to learner language than to L1-speech and, therefore, is often paired with the concept of accuracy (Chambers 1997: 536; see also Brand and Götz 2011). Viewed as a concept intertwined with accuracy, the approach of applying fluency to L1-speech may seem absurd. Fluency understood as a marker for language proficiency is not applicable to L1-speakers, which is why Riggenbach observes that “in common usage rarely does one hear

a native speaker being called fluent in comparison to other native speakers” (1991: 424). If someone’s first language is, for example, British English, speakers are simply considered as being fluent in English and the same perception is true for speakers of other L1-varieties of English.

However, if one’s view of fluency is broadened and detached from the concept of accuracy, it can also be used to describe and analyse L1-speakers’ utterances. Producing speech always necessitates bridging the gap between thought and perceivable language. Given the rapid pace of oral speech production, this poses a challenge even for highly proficient speakers. Segalowitz labels this as “utterance fluency,” that is, “the features of utterances that reflect the speaker’s cognitive fluency” (2010: 165). Utterance fluency can be operationalised as the total or relative amount of fluencemes a speaker uses. Accordingly, in our understanding, we do not consider such strategies negatively connoted markers of disfluency (e.g. Maclay and Osgood 1959), but rather as fluency-enhancing strategies.

2.1. Utterance fluency in L1-varieties of English

There are some corpus-based studies on fluency in L1-varieties of English such as Götz (2013), Osborne (2013), Crible (2018), and Dumont (2018). Although these studies’ set-ups slightly differ from ours with, for example, Crible (2018) and Dumont (2018) including conjunctions and Osborne’s (2013) focus lying on temporal fluency and syntactic patterns, their findings are still relevant for our study. Crible, for example, reports that “fluencemes are omnipresent in speech production, covering about one fifth of the sound signal, and such momentary interruptions of the smooth unfolding of speech mostly attend to the upcoming rather than the previous material” (2018: 133). The sheer rate of fluencemes found in Crible’s data already indicates that they “are the normal accompaniment of [...] speech” (Dumont 2018: 63). Fluencemes require further research into how they are distributed and what factors influence them. For this perspective, it is useful to take a closer look at the different categories of fluencemes that will be analysed in Section 4. When describing such fluency-enhancing devices, there are different labels to choose from. In this paper, we will adopt the terms ‘filled pauses’, ‘unfilled pauses’ and ‘discourse markers’ to refer to the three fluencemes we investigate, which are the three most frequently occurring strategies in speech, and thus we consider them to be ‘core fluencemes’.

2.2. Previous research on fluencemes in L1-English

Discourse markers have taken centre stage in previous research on L1-English (e.g. Brinton 1996; Aijmer 2002; Beeching 2016; Crible 2018, to name but a few). There has been extensive work on the social and pragmatic functions of discourse markers. Generally, there is consensus that discourse markers have no effect on the truth condition of an utterance, but Fung and Carter (2007: 414) point out that omitting them also means omitting clues for the listener as to how the truth condition is to be interpreted. Many different functions have been attested for discourse markers, for example, acting as planners, floorholders, and repair markers (Aijmer 2002: 51), as well as facilitating ease of listening for the reader (Aijmer 2002: 3), or as discourse initiators, boundary markers, floor managers, indicators of information status or to make implicit information explicit, functioning as responses to previous discourse or as hedging devices (see Brinton 1996: 37–38), as well as marking boundaries of talk and ending topics (see Fung and Carter 2007: 412). For the purposes of the present study, however, their varying functions beyond planning are of somewhat secondary importance, although discourse markers are not restricted to one discourse function alone and can potentially fulfil several of their attested functions at the same time. Beeching (2016: 99), for example, argues that the discourse marker *you know* is used for hesitation while being directed at the listener and asking the interlocutor to collaborate in bridging the planning phase.

Specific functions aside, discourse markers have been reported to “vary in adult native varieties depending on social context” (Fuller 2003: 192). In New Zealand English, for example, social class has been identified as a determining factor for the use of discourse markers such as *eh* and *you know*, which have been identified as features of the vernacular (see Stubbe and Holmes 1995: 74). However, the same study also concludes that New Zealand English shows a comparably small class effect in contrast with other varieties (1995: 72–73).

In addition to this kind of variety-internal variation, there are also differences in discourse marker use across different varieties. Speakers of New Zealand English use the discourse marker *like* significantly less frequently than their Australian counterparts, at least in private conversations. In Australian data, the use of *like* varies by the speaker’s age: while teenagers use *like* frequently, a decrease in frequency can be observed when speakers approach adulthood. This effect, however, cannot be found in New Zealand English where *like* is not age-restricted (see Miller 2009: 327–328), indicating an age-by-

variety interaction. In a similar vein, Reichelt (2021) shows that discourse markers can also vary across social factors. She finds that the choice between the discourse markers *kind of* and *sort of* is sensitive to age, amongst other factors. Other sociolinguistic variables have also been attested to influence discourse marker usage (e.g. Tagliamonte 2005 on gender; Fuller 2003 on speech context). In addition, Buysse and Blanchard (2022) demonstrate that L1-English speakers have an overall positive attitude towards discourse markers, particularly associating them with friendliness and confidence. In fact, even though discourse markers tend to be marked as signals of a low education level and lack of politeness, the L1-speakers almost invariably rated speech with a frequency of discourse markers much higher than their L2 counterparts (see Buysse and Blanchard 2022: 239–240), suggesting that discourse markers are an accepted asset to L1-speech. Thus, there is a solid foundation of research showing that discourse markers are inextricably linked to their social context. Although we only focus on the planning function of discourse markers, we also expect to see connections to sociolinguistic variables when we take a comparative approach to fluency across varieties of English.

The second most frequently researched fluenceme is filled pauses, such as *er/uh* or *erm/urm*. Filled pauses' prolific research history seems natural, given that they are reported to make up 6.38 per cent of speech (see Clark and Fox Tree 2002: 81). Filled pauses are particularly important in fluency research, because they are closely connected to a speaker's utterance fluency (e.g. Clark and Fox Tree 2002; O'Connell and Kowal 2005, among many others). Previous research has attested that filled pauses can be considered as words or 'true' linguistic items that follow clear usage patterns (e.g. Kjellmer 2003; Tottie 2011, 2015). Kirjavainen *et al.* (2022) support the classification of filled pauses as grammatical items. Their findings suggest that, at least in British English, the placement of filled pauses is relatively constrained. They also find that listeners typically do not notice the presence of filled pauses, similar to grammatical structures. Consequently, they tentatively suggest that speakers could be processing filled pauses in the same manner as grammatical structures (Kirjavainen *et al.* 2022).

A large body of research into filled pauses was contributed by Gunnel Tottie. Her work on filled pauses ('planners' in her terminology; Tottie 2011: 193) includes contrastive findings on two L1-varieties of English, namely British and American English. She finds that filled pauses display a higher frequency in British than in American English, and she also attests differences in other usage patterns, for example,

in the co-occurrence with unfilled pauses and the distribution of nasalised and non-nasalised variants of filled pauses (Tottie 2015: 42, 47, 49). Tottie (2011) also identifies several sociolinguistic factors as significant predictors of filled pause usage. In her data, men produce more filled pauses than women (2011: 182). Additionally, filled pauses are most frequent among people who are 60 years or older, although the data do not clearly indicate an overall age-grading effect, given that there is no discernible trend in other age groups (2011: 186–187). Lastly, her data show social class to play a significant role. Speakers from what she identifies as the highest socio-economic class produce significantly more filled pauses than other speakers (2011: 188).

Although there is some debate about the terminology of filled pauses and their exact functions (see, e.g., Revis and Bernaisch 2020: 137–138), with some perception studies even suggesting that filled pauses are used more frequently when speakers are lying (e.g. Loy *et al.* 2017), they occur relatively frequently in their function as fluencemes (Maclay and Osgood 1959: 34; Götz 2013). Their other functions notwithstanding, it stands to reason that filled pauses in particular count towards phenomena that are used systematically to overcome planning phases. O’Connell and Kowal arrive at the same conclusion and assert that *uh* and *uhm* “protect and sustain genuine fluency” (2005: 527).

The third core fluenceme we are concerned with in the context of the present study is unfilled pauses. Their occurrence in speech is considered to be a natural (Biber *et al.* 1999: 1054) and non-random phenomenon (see Svartvik 1990: 73). While some unfilled pauses are placed intentionally by the speaker to fulfil clear communicative functions (Drommel 1980) and are perceived as being natural by the listener (Chafe 1980), others occur during breathing or gesturing, etc. However, the majority of unfilled pauses occur for speech-planning reasons (Drommel 1980). They have been shown to occur most frequently at major grammatical transition points, for example, between syntactic units or at points where utterance launchers (such as *oh*, *well*, *okay*) are likely to occur (see Biber *et al.* 1999: 1054). In a similar vein, Lounsbury (1954) distinguishes between what he calls ‘juncture pauses’ at major syntactic boundaries, and ‘hesitation pauses’ within syntactic units, where listeners are more tolerant to the former and perceive the latter as being longer (Butcher 1980). Pausing within constituents and towards clause boundaries has been shown to be quite common and natural in speech (e.g. Pawley and Syder 1983: 200; Götz 2013; Revis and Bernaisch 2020: 139). Maclay and Osgood (1959: 31) have found that L1-English speakers tend to use unfilled pauses before beginning a phrase and

then before uttering the content word in that phrase. This pattern of placing pauses at mid- and end-clause position within an utterance is one feature of what is labelled “breakdown fluency” by Foster and Tavakoli (2009: 875). Unfilled pauses occur with such high frequencies in their function as fluencemes that they can be seen as good indicators of a person’s overall productive/utterance fluency (see Goldman-Eisler 1961; Segalowitz 2010). That being said, not much is known about unfilled pauses and their usage in correlation with sociolinguistic variables.

As our literature review reveals, most research conducted so far presents itself in the form of (case) studies of small sets of fluency-enhancing devices of individual fluencemes on individual discourse markers. Studies covering different varieties include Burrige and Florey (2002), Miller (2009), Burrige (2014), Vine (2016), and Burke and Burrige (2023) on discourse markers in Australian and New Zealand English respectively, as well as Grant’s (2010) comparison of *I don’t know* in British and New Zealand English. Canadian English appears to be rather under-researched in terms of fluency, although there has been research on individual discourse markers, for instance, Tagliamonte’s (2005) study on the use of *so*, *like* and *just*. These comparative studies indicate that there is some degree of variation between fluencemes across different varieties of English. It has also been shown that sociolinguistic variables can influence fluenceme use (e.g. Bortfeld *et al.* 2001 for American English). We therefore assume to find differences in fluenceme use to be caused by (an interplay of) different varieties of English and sociolinguistic variables, such as gender, age or occupation.

2.3. *Towards a contrastive approach to fluency across L1 varieties of English*

The paper at hand is concerned with British English (BrE), Australian English (AusE), New Zealand English (NZE), and Canadian English (CanE).² We chose to work with these four L1-varieties as we wanted to compare differences between varieties spoken on different continents and in countries in which they constitute one of the dominant languages or even the dominant or official language (Kachru 1985: 12). Because these

² While it would have been desirable to include American English in our analysis, we opted for Canadian English instead, because it is the geographically closest neighbour to American English for which we had the same type of data as for the other varieties. Since fluency has been shown to be very register- and text-type sensitive, we wanted to make sure to only compare data that were as similar as possible.

varieties have been extensively researched with respect to regional or social variation within each variety itself, we will not describe them further at this stage.

All these four Englishes under scrutiny but BrE are severely under-researched in terms of fluency in general and in terms of comparative fluenceme use in particular. Against this backdrop, this paper takes a first step toward addressing this research gap by examining potential differences in fluenceme use across four varieties of English. We also incorporate sociolinguistic factors into our analysis to explore their role in determining the choice of specific strategies. We also test for interactions between these variables as it might be the case, for example, that gender affects fluency performance, but only for speakers of a certain variety.

3. DATABASE, CODING PROCEDURE, AND APPLIED METHODOLOGY

3.1. Database and coding procedure

We selected four components of the *International Corpus of English* (ICE), namely the BrE (ICE-GB), AusE (ICE-AUS), NZE (ICE-NZ), and CanE (ICE-CAN) components. The ICE corpora are highly suitable for our purposes, as all ICE components were compiled using the same design and applying common schemes for annotation, which ensures comparability across different sub-corpora (Greenbaum and Nelson 1996).³ Each component contains 300 spoken text files of approximately 2,000 words per file from different types of settings, such as dialogues and monologues, private, public, scripted, and unscripted speech. We chose 100 texts containing unscripted private dialogues from either face-to-face or telephone conversations, because we consider these data to be the most natural data type in the spoken section of the corpus.⁴ Our database thus comprises *ca.* 200,000 words per variety and *ca.* 800,000 words in total. Managing and coding such a large amount of data was possible by taking several steps. We initially identified relevant fluencemes by carefully reviewing the pertinent literature and manually adding further items to the categories under investigation (i.e., unfilled pauses, filled pauses, and discourse markers). For each subcorpus, we manually added variety-specific fluencemes.

³ Of course, there are some remaining issues with regard to the comparability of the different components of the ICE components, but we consider them to strike the best balance in terms of comparability and diversity of registers/genres represented in the varieties' components.

⁴ Obviously, the analysis proposed here is based on the assumption that the corpus annotation is sufficiently precise and attempts to benefit from the fact that, while certainly less precise than acoustic measurements, it at least involves many more speakers than such analyses can reasonably be expected to include.

For example, in ICE-NZ we found that the items *yeah no* and *yeah nah* are used as fluency enhancing devices (see also Burrige and Florey 2002; Manhire 2021). We added such kinds of frequently occurring discourse markers so as not to miss variety-specific fluencemes which might be used as alternative strategies to other discourse markers.

The identified fluencemes were uploaded to a coding application developed by Wolk *et al.* (2021) in *R* (R Core Team 2024) utilising the web application framework *shiny*.⁵ This tool is able to identify the uploaded items via string search in the text files of the corpora under investigation. The tool's interface displays the transcribed conversations, which are only available as unformatted text files, as dialogues and highlights the identified items. As the tool identifies strings and not syntactic categories, it is necessary to manually disambiguate the detected items. This is illustrated in the examples in (1). In (1a), *like* is a verb, not a fluenceme, but in (1b) *like* can be considered to be a fluency-enhancing device.

(1a) I *like* his new car.

(1b) This is *like* really new information.

Given the extreme variability of spoken language, the sometimes-erroneous transcriptions, and missing information on intonation, the process of manual disambiguation is extremely challenging, so we took several measures to ensure reliable annotation. A coding manual was issued and constantly revised and adapted during the coding process and we trained our coders using training files and coded every file in several steps. First, each file was initially coded by a student assistant. A second student assistant reviewed the entire file for coding accuracy, noting any discrepancies and discussing them with the original coder. Items on which agreement could not be reached were referred to a senior researcher for resolution. If the senior researcher was also uncertain about the item's status, it was brought before the entire team of four student assistants and three senior researchers. A majority vote was used to decide if consensus could not be reached. Truly ambiguous cases were deleted from the dataset. We deemed this four-step coding process superior to a double-blind coding approach that merely calculated annotator deviations, as it allowed us to carefully discuss the contexts of any potentially ambiguous fluencemes. This ensured that our dataset consisted solely of truly

⁵ More information about *shiny* is available at <https://shiny.rstudio.com>.

accurate hits. The coding guidelines we use will be summarised in the following with reference to exemplary problems that occurred during the annotation process.

3.1.1. Discourse markers

Inspired by constituency tests which can be found in most syntax textbooks (e.g. Börjars and Burridge 2010; van Gelderen 2010), we consider discourse markers as fluencemes if (i) they can be omitted without either changing the truth value of an utterance, (ii) if they can be omitted without rendering the utterance ungrammatical, or (iii) if they can be replaced by another fluenceme or another discourse marker. Although these simple tests worked for a large part of the dataset, numerous instances remained that were difficult to code. One example is potential fluencemes that occur in utterance-final position. Theoretically, an utterance-final *like* can more often than not be a fluenceme or also fulfil another syntactic function, as shown in example (2).

- (2) Speaker A: I started university this year. It is *like*...
 Speaker B: What do you study?

As Speaker A was interrupted by Speaker B, it is impossible to tell what exactly Speaker A wanted to say. For example, it could be *It is like school*, in which *like* does not act as a fluenceme, or *It is like really boring*, where it has clear discourse marker status. Thus, we chose to not include such pre-interruption items, as well as all other cases that were truly ambiguous. Altogether, we annotated and disambiguated 33 different types of discourse markers and smallwords that were used in a planning function, as illustrated in Table 1.

3.1.2. Filled and unfilled pauses

Filled pauses function as fluencemes in most cases, but still needed to be disambiguated. Speakers especially in NZE and AusE, for example, also frequently use these sounds as question tags (e.g. *You saw that, eh?*), either to react to an utterance or as a backchanneling response (e.g. Schweinberger 2018). Since such instances do not serve the fluency-enhancing function we are investigating, they were excluded from our dataset, as well as all other functions in which the filled pauses listed were not used as planning devices (e.g. backchannels, emphasisers, etc.). If it was impossible to determine their function, they

were deleted from the dataset. The set of filled pauses analysed in our study consists of 19 types, which are illustrated in Table 1.

Fluenceme	Subtypes
Discourse markers	<i>actually, alright, alright, anyhow, anyway, basically, do you know what I mean, huh, I don't know, I mean, in a way, kind of, kinda, know what I mean, let's see, like, nah, no, okay, or something, right, so, sort of, sorta, stuff like that, thing like that, things like that, well, yeah, yep, yes, you know, you see</i>
Filled pauses	<i>ah, ahh, ahm, ahn, eh, ehm, er, haan, hmm, mm, mmm, mhm, uhm, uhn, hm, ur, urm, uh, uhh</i>
Unfilled pauses	<i><, >, <., ></i>

Table 1: Overview of investigated types of fluencemes

Finally, unfilled pauses are automatically accepted as fluencemes by the corpus tool and are not manually disambiguated.

3.2. Clean-up and annotation of the dataset

Following the disambiguation, the resulting fluenceme data were further cleaned. Crucially, as we study L2-English speakers, all speakers who had indicated a first language other than English were removed from the dataset. As the central point of the analysis, we then conflated the fluenceme categories. In the coding tool's initial output, each fluenceme is categorised in the fine-grained manner detailed in Table 1 above and Table 2 below so that, for example, one can distinguish between the different realisations of discourse markers. For the present analysis, we adopted a broader three-way categorisation with the levels 'discourse marker' – 'filled pause' – 'unfilled pause'. Next, we computed individual fluenceme counts for each speaker with attested fluenceme usage in the data. To this end, we used the corpus tags to automatically identify in each file any non-corpus material (e.g. insertions from the transcribers), which was then removed in order to compute the number of words uttered by each speaker. Using these word counts per speaker per file, we computed the response variable, that is, three normalised frequencies of fluencemes (one per fluenceme type) for each speaker, by dividing the absolute frequency of each fluenceme type by the number of words per speaker. Due to the nature of this normalisation, we were unfortunately not able to include the linguistic context of the fluencemes (e.g. the frequency or complexity of the lexeme following a fluenceme, or whether or not it was primed) to our analysis, although we believe that adding such 'level-1' variables (see Gries 2024) at a later stage could add more explanatory power to our analysis.

Lastly, we added (speaker-specific) information from the corpora’s metadata to be included in the analysis. Firstly, we entered both the variable CORPUS and the text category variable TEXTCAT for each speaker, that is, whether their data come from a face-to-face conversation or a phone call. Next, we annotated the speakers’ GENDER (i.e. *male* or *female*) as specified in the metadata. The metadata also included information on the speakers’ AGE, but with different binning between the four corpora. To account for this, we kept all individual ages (i.e. non-binned ones) as is and transformed all age ranges to their median. All four sets of metadata also provided information on speaker education and speaker occupation, which we also needed to homogenise, because the information given was too diverse (i.e. 94 different levels for education and 242 levels for occupation). We therefore used the format inspired by the scaling of indices related to the Pisa index of economic, social and cultural status (OECD 1999). For the speakers’ educational level EDUCATION_COARSE, we applied the format of the International Standard Classification of Education ISCED-97 (UNESCO 2006) and transformed the speakers’ education to a Level from 0–6, where 0 corresponds to elementary school and 6 to postgraduate education, such as a PhD (OECD 1999). Each occupation, as stated by the speakers in the corpus, was entered into a variable OCCUPATION_ISCO880 transformed to the value corresponding to the International Standard Classification of Occupations (ISCO-08 (COM), ranging from 16 (e.g. farmers) to 85 (e.g. medical professionals/jurists) (Ganzeboom 2010).

Lastly, we added each SPEAKER identifiable by a unique ID as a variable to the dataset. This led to us having 51,950 data points for which all of the above-listed predictors available distributed across 719 speakers, as summarised in Table 2.⁶

	AustrE	CanE	BrE	NZE	Sum
Discourse markers	4,854	6,368	3,745	101	15,068
Filled pauses	1,966	3,444	2,713	48	8,171
Unfilled pauses	11,012	10,771	6,884	44	28,711
Sum (# of speakers)	17,832 (197)	20,583 (248)	13,342 (265)	193 (9)	51,950 (719)

Table 2: Overview of the final data

⁶ We are not providing a more detailed descriptive statistics breakdown per fluenceme and/or variety because these cannot do justice to the multifactorial nature of the data: mean frequencies, etc. grouped by only one predictor *per definitionem* come with a huge loss of information because they are aggregated over all other predictors.

3.3. Statistical analysis

Two kinds of analyses are possible for this kind of data, which come with different implementations:

- a type 1 analysis, in which the response variable is the fluenceme type (i.e., ‘discourse marker’ vs. ‘filled pause’ vs. ‘unfilled pause’) and where one would try to predict from the other variables what disfluency will be chosen in every single utterance case/context;
- a type 2 analysis, in which the response variable is the fluenceme type frequency (i.e. a numeric value of 0) and where one would try to predict the frequency with which speaker groups as defined by the other variables use which fluenceme how often.

Type 1 analyses are more common in corpus studies of variationist phenomena and might seem like the natural choice here. However, most such studies focus on hypothesis testing and include level-1 predictors (Gries 2024) that vary with the linguistic choice being studied. In contrast, our exploratory study aligns better with a kind of type 2 analysis focusing on how speaker groups, defined by annotated variables, tend to favour certain fluencemes. Thus, we adopt type 2 analysis here, reserving type 1 for future research. Type 2 analyses with response variables that are frequencies can be analysed in many ways, including various kinds of regression models, but we chose the tree-based method of random forests, which can be applied to frequencies without link functions, does not make any assumptions about the data,⁷ can handle potentially collinear predictors slightly better than regressions, and is extremely good at detecting interactions, that is, “identifying ‘the interaction effect per se and the predictor variables interacting with each other’” (Gries 2021: 481). The random forests approach is based on fitting many different classification or regression trees on a data set (e.g. 500), but adding two layers of randomness:

- each of the trees in the forest is fitted on a different randomly sampled-with-replacement version of the data;

⁷ Even count-specific regression models have assumptions (Poisson regression requires equal expectation/mean and variance, while negative binomial regression needs a link parameter) that tree-based methods, including forests, do not have.

- in each tree, only a randomly-selected subset of the predictors is eligible to be chosen for a split (see Gries 2021: 463–464).

These two characteristics are desirable inasmuch as they help to reduce the risk of overfitting (because trees are fit on many different versions of the data and because the prediction accuracy is evaluated against the data that did not make it into a specific sample) and in how they decorrelate trees (e.g. as very powerful predictors are less able to dominate every single tree because they might not always be available for a split). In addition, random forests (i) can be less sensitive to distributions that pose problems for regression, (ii) are good for small n -large p problems (i.e. scenarios with many predictors but not many data points), and (iii) often outperform regressions when it comes to making accurate predictions (e.g. because they can be more powerful at detecting non-linearity).

The main disadvantage of random forests is the challenge they pose for interpretation: unlike in a regression, there are no simple coefficients for differences between means or slopes, and unlike in a tree, there is no simple visualization of potentially thousands of trees in a forest. In addition, detecting interactions (i.e. identifying the interaction effect of 2+ predictors) is not as straightforward as is widely believed (see Wright *et al.* 2016; Gries 2020, 2021: 481–483). To shed light on the inner workings of the fairly black-box random forest, we use

- **variable importance scores** (permutation-based); note, these indicate “the impact of each predictor variable individually as well as in multivariate interactions with other predictor variables” (Strobl *et al.* 2009: 337);
- **partial dependence scores**; these indicate what in a regression model would be the direction/sign of effects of predictor values/levels. Note that, to identify interactions between predictors, we computed partial dependence scores not just for predictors on their own, but also for selected interactions of interest.

Even though random forests are already very powerful and flexible by default (the trees they are based on do not make distributional assumptions), they sometimes can benefit from at least some minor degree of preparation. In our case, the normalised frequencies of each fluenceme per speaker were very right-skewed. While the values ranged from 0 to 0.25, more than 80 per cent of them were below 0.05, so we applied a Box-Cox

transformation to avoid a long underpopulated tail on the right;⁸ as a result, the new version of these normalised frequencies was very symmetric.

The random forest implementation we chose for this study is the one implemented in the *R* package `{ranger}` (Wright and Ziegler 2017), which also provides the variable importance scores; the partial dependence scores were computed with the *R* package `{pdp}` (Greenwell 2017) and we followed a logic similar to the one discussed in Gries (2021: 480–483). The formula submitted to `ranger::ranger` was the one shown below and we built `ntree=3,000` trees with the default settings of `ranger`'s hyperparameters `mtry=2`, a target node size of 5, and permutation-based local variable importance scores.

```
NORMFREQ_bnc ~
  FLUTYPE + # discourse marker vs. filled pause vs. unfilled pause
  CORPUS + # Austr vs. Canadian vs. British vs. New Zealand
           English
  TEXTCAT + # direct vs. distance conversations
  SPEAKER + # the unique ID of each speaker (≅ a random effect)
            # and several speaker-related variables:
  AGE + GENDER + EDUCATION_COARSE + OCCUPATION_ISC088
```

4. RESULTS

Given that the response variable was only a normalised frequency and, thus, disregarded any contextual level-1 variables (see Gries 2024), the out-of-bag prediction accuracy was surprisingly high with an R^2 of 0.492. Figure 1 shows the relationship between the observed and predicted fluenceme frequencies with the modelled transformed frequencies on the bottom and left axes and the back-transformed frequencies on the upper and right axes.

⁸ We used the function `car::bcnPower` (Fox and Weisberg 2019) to estimate the best lambda (-2.094296) and the best gamma (0.1) values for the response variable.

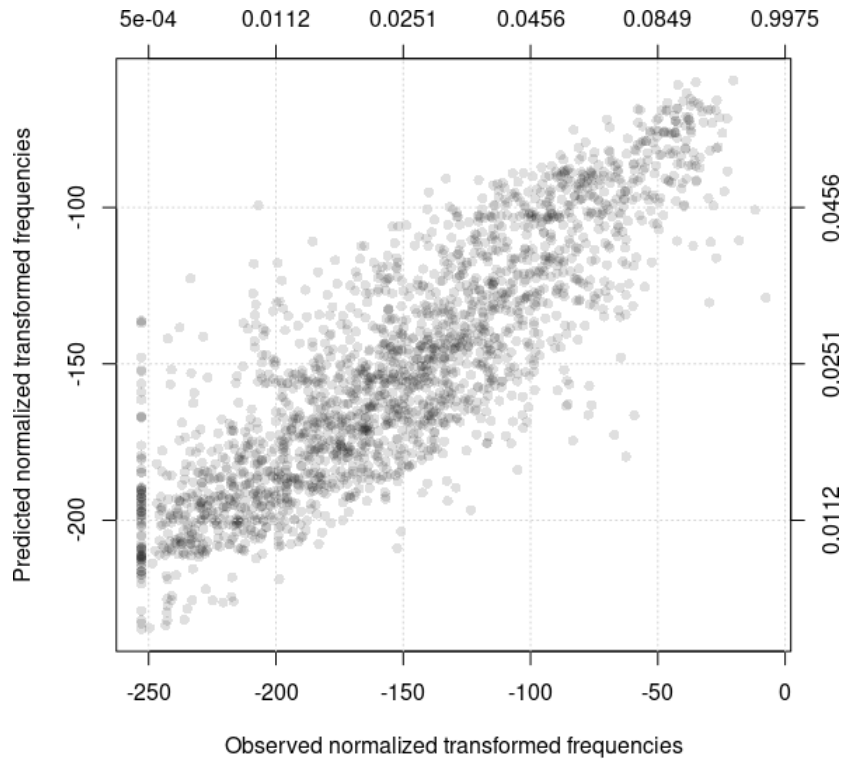


Figure 1: Predicted vs. observed proportions of fluenceme frequency

The variable importance scores returned the following order of predictors: FLUTYPE (2,519) >> SPEAKER (648) > CORPUS (469) > AGE (266) > OCCUPATION_ISCO88 (199) > EDUCATION_COARSE (78) \approx TEXTCAT (73) \approx GENDER (67). But the more interesting questions are of course how the most important predictors affect the prediction of the response—alone (as main effects) or in interaction patterns—and in which direction they do so. We therefore computed partial dependence scores (back-transformed to the original proportions for interpretability) for the predictors FLUTYPE, SPEAKER, CORPUS, and EDUCATION_COARSE, as well as for the ‘interactions’ FLUTYPE:CORPUS and AGE:GENDER(:FLUTYPE).

4.1. The effects of FLUTYPE, CORPUS and FLUTYPE:CORPUS

As might be expected from the importance scores, the effect of the predictor FLUTYPE is fairly strong: unfilled pauses are by far the most frequent (0.04), followed by a large margin by discourse markers (0.023), and filled pauses (0.014), as shown in Figure 2.⁹

⁹ While some researchers might object to the use of lines connecting the dots (given that, indeed, there are no data between, say, discourse markers and filled pauses), we still prefer using lines because of how they help guide readers’ visual processing of the relations between data points (see also Fox and Weisberg’s 2019 effects package).

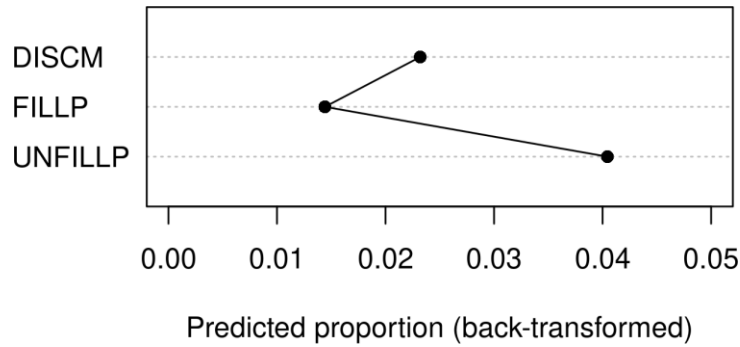


Figure 2: Predicted proportions of fluencemes by FLUTYPE

The effect of CORPUS is somewhat weaker and is largely driven by how NZE speakers behave differently from the rest: Australian, Canadian, and British speakers are quite similar (0.026, 0.026, and 0.024, respectively) whereas the NZ speakers have a notably lower frequency of 0.016 (see Figure 3). However, this effect of CORPUS must not be overinterpreted because (i) NZE was only represented by nine speakers in our data, and (ii) all of those were from one education level (the second highest).¹⁰

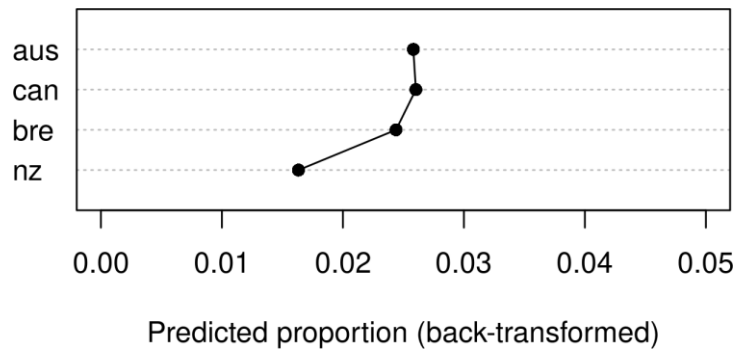


Figure 3: Predicted proportions of fluencemes per CORPUS

Given our interests, it was worthwhile to determine whether (part of) the reason for the high variable importance scores was an interaction of these two predictors. We therefore computed partial dependence scores for the interaction, both perspectives of which—in terms of which predictor is grouped by which other one—are shown in Figure 4.

¹⁰ The degree to which the NZE speakers are different from the speakers in the other varieties might lead some to suggest leaving them out, but we would consider that a mistake. We think it is better to provide the (limited) descriptive information they afford in the forest together with the context required to (not over)interpret the results than it is to simply pretend those data do not exist. For example, now that we have described the data to at least some extent, follow-up studies can use our description as a launchpad to motivate further data collection or to drill down into only the highest education levels of the other varieties.

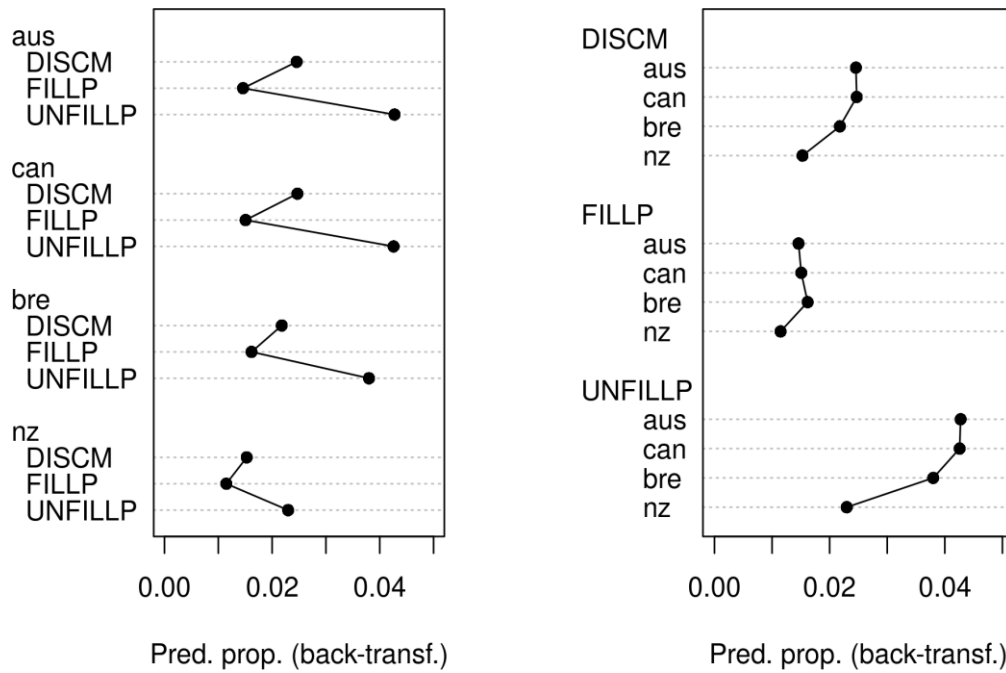


Figure 4: Predicted proportions of fluencemes for FLUTYPE:CORPUS

It seems that there is only a weak interaction. In all four varieties, unfilled pauses are much more frequent than discourse markers, which are more frequent (but less so) than filled pauses. However, NZE is peculiar because it features a much smaller number of unfilled pauses than any other variety. From a different perspective, we can see that

- the frequencies of discourse markers and unfilled pauses across varieties behave similarly to the frequencies of all fluenceme types across varieties;
- the frequencies of filled pauses vary least across varieties.

4.2. The effect of *SPEAKER*

There is a great degree of speaker heterogeneity, but given the high number of speakers and the lack of interpretability of individual speaker preferences, this can only be shown heuristically, as we do in Figure 5.

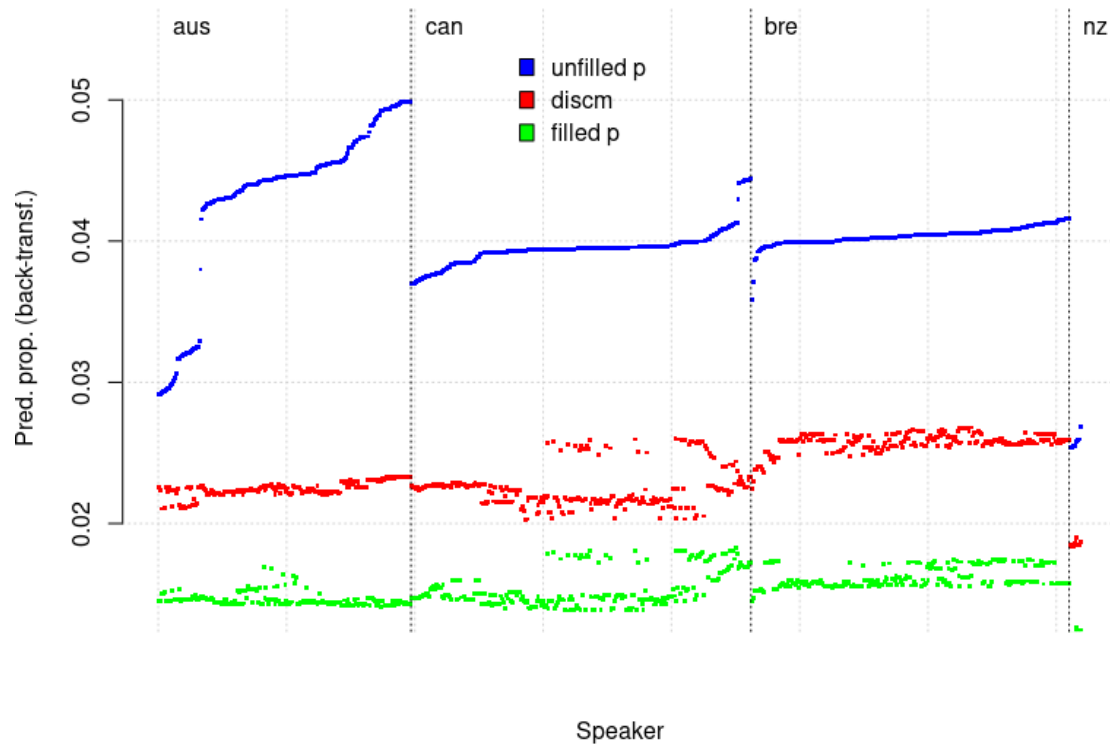


Figure 5: Predicted proportions of fluencemes per FLUTYPE:VARIETY/SPEAKER

The y-axis represents the predicted proportions of fluencemes with each fluenceme type in a different colour and the x-axis representing all 719 speakers grouped by variety. Note that the x-axis should not be interpreted sequentially, as it is not a variable like TIME; instead, one must interpret it vertically such that each position on the x-axis represents one speaker's three fluenceme frequency predictions. Most visually striking is of course the huge variability we observe for the unfilled pauses (in particular for the AusE speakers), but it is also worth noting that, for discourse markers and filled pauses, variation coefficients¹¹ indicate that the BrE speakers are most variable.

4.3. The effects of AGE, GENDER, AGE:GENDER, and AGE:GENDER:FLUTYPE

The two variables AGE and GENDER do not have particularly strong effects, but we were ultimately also interested in determining (i) whether the genders differed in their overall fluenceme differences, (ii) whether there was an age effect on overall fluenceme use, and (iii) whether any of these effects interacted. While this was not particularly likely, given the low variable importance scores of AGE and GENDER, we therefore computed partial

¹¹ For unfilled pauses (in blue), we obtained AusE (0.1337) >>> BrE (0.0322) > NZE (0.0176) ≈ CanE (0.0154); for discourse markers (in red), the order and values of the variation coefficients are: BrE (0.0619) >> CanE (0.025) ≈ AusE (0.0227) > NZE (0.0104); for filled pauses (in green), we obtained BrE (0.0823) >> CanE (0.0447) > AusE (0.0325) > NZE (0.0128).

dependence scores for what corresponds to a three-way interaction, which is represented in Figure 6.

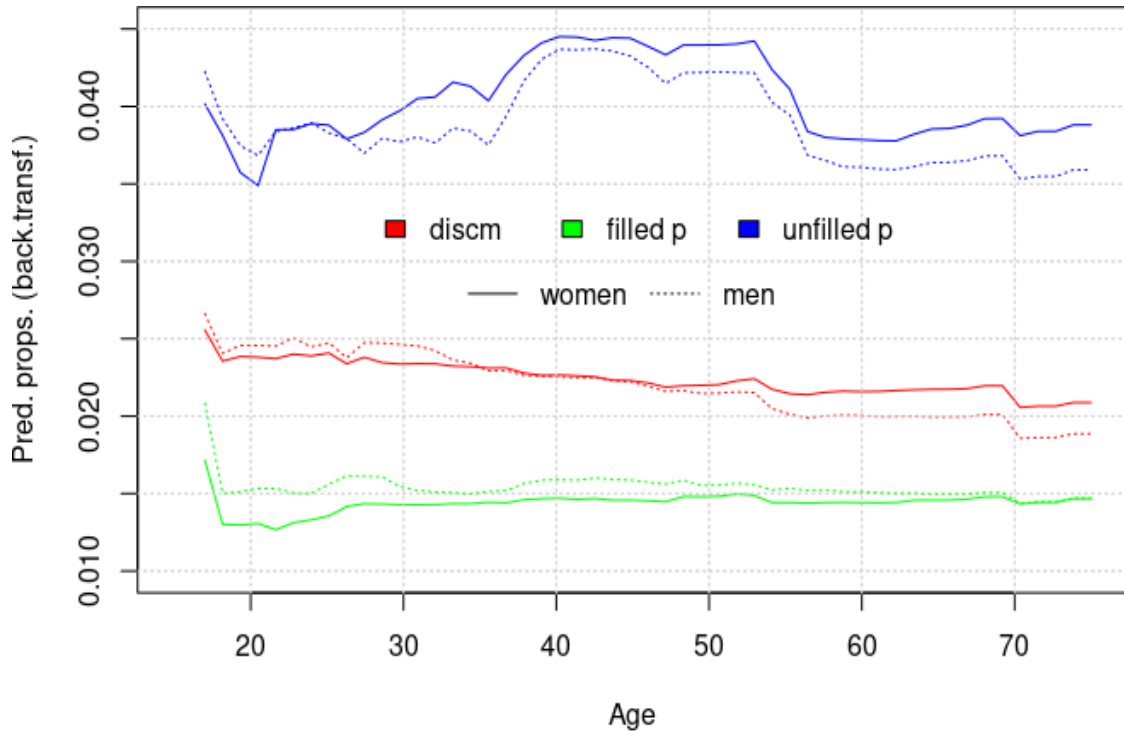


Figure 6: Predicted proportions of fluencemes for AGE:GENDER:FLUTYPE

The overall effect of FLUTYPE is the one we already know from above, and the overall effect of GENDER is indeed tiny, but there are at least some small tendential differences in how AGE is correlated with the response; for instance,

- unfilled pauses are most frequent for middle-aged speakers (both men and women);
- for discourse markers, there is a negative correlation with AGE, which might be ever so slightly stronger for men;
- filled pauses are most frequent for the youngest speakers, but largely remain uncorrelated with AGE once speakers reach 30 years of age.

4.4. The effect of EDUCATION_COARSE

The effect of EDUCATION_COARSE is weak and only slightly suggestive. Figure 7 appears to indicate that

- filled pauses are mildly negatively correlated with degree of education;

- unfilled pauses seem very weakly negatively correlated with degree of education;
- discourse markers are very weakly positively correlated with degree of education.

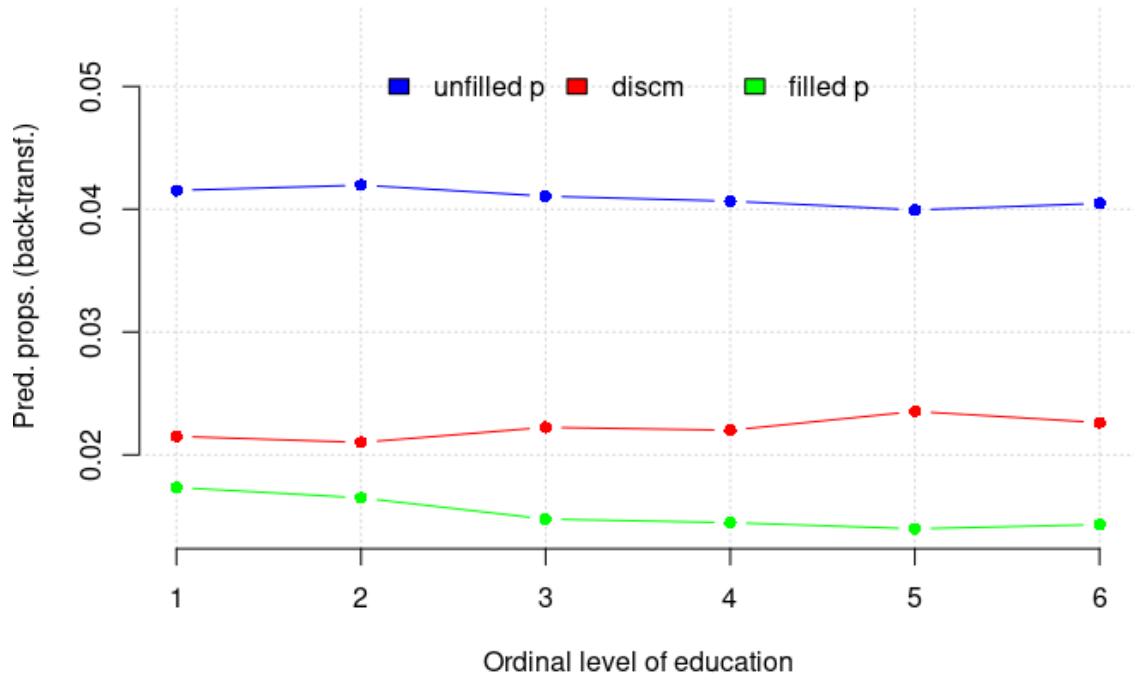


Figure 7: Predicted proportions of fluencemes for EDUCATION_COARSE

5. DISCUSSION, CONCLUSION AND OUTLOOK

This study explored the number and distribution of fluencemes across four L1-varieties of English, accounting for the influence of sociolinguistic variables. To our knowledge, it is the first study to investigate different fluencemes in combination across several L1-varieties of English. Looking at fluency in this holistic way enabled us to uncover complex and interactional effects between the different variables. We will discuss these findings in relation to our research questions in the following.

Regarding our first research question, that is, whether there are differences in fluenceme use between various English varieties, the answer is both affirmative and negative. While the variety of English indeed made a difference, the distribution of fluencemes across all four varieties remains similar. Notably, NZ English speakers generally use fewer fluencemes, particularly fewer unfilled pauses but, as we cautioned above, this could be due to the small number of NZ English speakers in our dataset or stem from potentially different transcription practices. Despite this, the overall

distribution of fluencemes is consistent across all varieties: unfilled pauses are used most frequently, followed by discourse markers, and then filled pauses. While it is reasonable to expect a similar degree of planning pressure among monolingual L1-English speakers, we had set out to test if individual varieties might exhibit unique preferences for specific types of fluencemes. However, the data suggest a striking similarity in terms of fluenceme frequencies, which might be due to their similar cognitive-functional role in speech planning. This points to an underlying “common [fluenceme] core” (Quirk *et al.* 1985: 16) across L1-English speakers. This core is

present in all varieties, so that, however esoteric a variety may be, it has running through it a set of [...] characteristics that are present in all the others. It is that fact that justifies the application of the name ‘English’ to all the varieties (Quirk *et al.* 1985: 16).

Given this comparable set of fluencemes across inner-circle English varieties, future studies should explore the specific realisations of fluencemes, that is, which types of discourse markers (e.g. *like* vs. *you know* vs. *I don’t know*) or which types of filled pauses (e.g. *uh* vs. *erm* vs. *ur*) along with potentially different phonological realisations are preferably used in certain varieties (e.g. Miller 2009; Tottie 2011, 2015; Fruehwald 2016), to determine if they qualify as *allo*-fluencemes. Such future investigations might yield more pronounced findings when examined in greater qualitative depth, such as by considering the position of the fluenceme within the utterance and the linguistic context following the fluenceme under scrutiny. Moreover, it would be particularly interesting to compare these findings with speakers of English as a second or foreign language. Understanding fluenceme preferences across different types of English varieties would significantly contribute to the discussion about a possible common fluenceme core.

As for the second research question —whether sociolinguistic variables influence the choice of particular fluencemes— the answer is only mildly affirmative. Contrary to our expectations based on previous research outlined in Section 2, the sociolinguistic variables we examined (gender, age, education, and occupation) exhibited much milder effects on speakers’ fluency than we had anticipated. This is particularly surprising in the case of age. Previous studies (e.g. Schow *et al.* 1978; Albert 1980) suggested that older individuals exhibit more planning phases than younger people, which we anticipated would correspond to a higher number of fluencemes in their speech. However, our findings did not strongly support this assertion, indicating that the influence of age on fluenceme use is subtler than previously thought. Instead, our findings revealed that the

highest number of unfilled pauses actually occurs in middle-aged speakers of both genders, while the number of filled pauses remains relatively stable from adolescence onwards for both genders. This result is puzzling, but the age-related effects, or lack thereof, are robust across varieties and a large number of speakers. Our findings thus seem to support the findings of a recent longitudinal study by Beier *et al.* (2023), which also found no significant increase in filled pauses among older individuals. Instead, their study suggested that older speakers tend to use different types of fluencemes, such as word repetitions, and exhibit different strategies to overcome planning phases, like slower speech. To obtain more reliable results regarding age-related fluency profiles, future studies should examine a broader range of fluenceme types across different English varieties. This would help clarify whether age-related patterns in fluency are consistent across various contexts and varieties.

The predicted age effect on discourse markers is also only mildly evident in our data, with their usage slightly decreasing as age increases. However, since our study did not distinguish between different types of discourse markers, this overall decline is relatively hard to interpret, since previous research indicates that some discourse markers increase with age, while others decrease or even remain constant (e.g. Miller 2009; Reichelt 2021). Given the vast variety of discourse markers, it seems necessary that future studies examine this variation more closely while controlling for sociolinguistic and geographic variables.

Turning to gender effects, gender also had an only mild and varied effect on the fluencemes we investigated. Whereas speakers from both genders show generally the same distributional patterns of fluenceme use, male speakers use more filled pauses than their female counterparts in all age groups (see also Shriberg 1996 or Bortfeld *et al.* 2001). Women use unfilled pauses more frequently than men from their early twenties onwards, a finding that we have not seen to have been discussed previously. As far as discourse markers are concerned, men use them more frequently until the age of approximately 35, when women's frequencies overtake those of men. These observations are only partly in line with Laserna *et al.*'s (2014) results, who find comparable frequencies of filled pauses across genders and ages, whereas discourse markers were more common among women in their dataset. Our findings add to the complexity of previous research output (e.g. Laserna *et al.* 2014; Scheuringer *et al.* 2017; Sokołowski *et al.* 2020). The gender effect on fluenceme use and its connection to utterance fluency therefore needs to be followed

up on further. One hypothesis in this vein has been put forward by Weiss *et al.* (2006), who suggest that men and women simply make use of different fluency-enhancing strategies to overcome planning phases, which at least partly also figures in our data.

However, although our findings do not clearly support specific gender- or age-related fluency profiles, we were still able to observe some overall trends, especially when looking at the three-way interaction between age, gender, and fluencemes. This emphasises the importance of controlling for such variables when investigating fluency. Potential follow-up studies could compare these findings with other fluenceme types to see if men/women or younger/older speakers make use of other strategies when they encounter planning phases (e.g. truncations, syllable-lengthening, repeats, etc.) or if generally different ways of easing the cognitive load of speaking can be observed, such as, for example, through a more frequent use of routinised formulae or even slower speech as in Götz's (2013) fluency groups or Dumont's (2018) (dis)fluency profiles.

Regarding the effect of speakers' educational level on their fluency, we observed only weak effects. There was a slight trend suggesting that more educated speakers use fewer filled and unfilled pauses (see also Stubbe and Holmes 1995), while an opposite, equally weak trend was observed for discourse markers. However, given the mild nature of these effects in our data, we do not consider ourselves in a position to draw any significant inferences from them at this point. Further research with a larger and more varied sample may be necessary to clarify the relationship between educational level and fluenceme use.

There are also some caveats to our study that we need to mention and which have the potential to turn into future research opportunities. For one, while our purely frequency-based approach to predicting fluencemes has revealed some interesting insights, the drawbacks have also become apparent. Especially for discourse markers and filled pauses, where we extracted a variety of different types, a purely frequency-based approach is extremely limited when it comes to interpreting the findings. It would therefore be particularly worthwhile to study the use of these different types along with their communicative functions and their utterance positions in a follow-up study in order to be able to reveal true variety-specific preferences of using certain discourse markers (in different functions or utterance positions) over others. On a more general level, this is especially relevant because at present our analysis does not include any linguistic and/or otherwise contextual level-1 predictors, that is, variables that describe context of an

individual choice for a particular fluenceme (see Gries 2024). For instance, it would be very interesting to determine whether there is a correlation between the (interaction of age and gender and) variety when it comes to investigating, for example which fluenceme is preferred depending on the complexity of the material directly following it.

Another phenomenon that falls into the same level-1 context category and that we have not been able to account for in the present study is the tendency of fluencemes to cluster together. For instance, discourse markers have been documented to frequently co-occur with filled or unfilled pauses or other discourse markers (e.g. Crible *et al.* 2017; Pons Bordería 2018). Therefore, looking at fluencemes in isolation only reveals limited information as for the number and quality of strategies speakers of different varieties of English require to fill planning phases. The same applies to utterance positions in which certain (clusters of) fluencemes occur, and even more so for the linguistic characteristics of a lexeme following a fluenceme (cluster), which we were not able to account for in the context of the present study. However, it would be extremely revealing to investigate in future studies.

Finally, we have to acknowledge that we have only investigated a limited number of only four inner-circle varieties of English. While this approach yielded promising results, there are contrastive research opportunities to take into consideration, such as how different variety types of English establish fluency, and whether the potential common fluenceme core can be extended to other types of Englishes.

REFERENCES

- Aijmer, Karin. 2002. *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins.
- Albert, Martin L. 1980. Language in normal and dementing elderly. In Lauraine K. Obler and Martin L. Albert eds. *Language and Communication in the Elderly*. Lexington, MA: DC Heath and Co, 145–150.
- Beeching, Kate. 2016. *Pragmatic Markers in British English: Meaning in Social Interaction*. Cambridge: Cambridge University Press.
- Beier, Eleonora J., Suphasiree Chantavarin and Fernanda Ferreira. 2023. Do disfluencies increase with age? Evidence from a sequential corpus study of disfluencies. *The Psychology of Aging* 38/3: 203–218.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Börjars, Kersti and Kate Burridge. 2010. *Introducing English Grammar* (second edition). London: Hodder Education.

- Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober and Susan E. Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role and gender. *Language and Speech* 44/2: 123–147.
- Brand, Christiane and Sandra Götz. 2011. Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics* 16/2: 255–275.
- Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin: Mouton de Gruyter.
- Burke, Isabelle and Kate Burridge. 2023. From *a bit of processed cheese* to *a bit of a car accident* and *a little bit of 'oh really'* – The journey of Australian English *a bit (of)*. *Journal of Pragmatics* 209: 15–30.
- Burridge, Kate. 2014. *Cos* – A new discourse marker for Australian English? *Australian Journal of Linguistics* 34/4: 524–548.
- Burridge, Kate and Margaret Florey. 2002. 'Yeah-no He's a Good Kid': A discourse analysis of *Yeah-no* in Australian English. *Australian Journal of Linguistics* 22/2: 149–171.
- Butcher, Alan. 1980. Pause and syntactic structure. In Hans-Werner Dechert and Manfred Raupach eds. *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton de Gruyter, 85–90.
- Buyse, Lieven and Meghan Blanchard. 2022. L1 and non-L1 perceptions of discourse markers in English. *Pragmatics & Cognition* 29/2: 222–245.
- Chafe, Wallace L. ed. 1980. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Chambers, Francine. 1997. What do we mean by fluency? *System* 25/4: 535–544.
- Clark, Herbert. H. and Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84: 73–111.
- Corley, Martin, Lucy J. MacGregor and David I. Donaldson. 2007. It's the way that you, *er*, say it: Hesitations in speech affect language comprehension. *Cognition* 105/3: 658–668.
- Crible, Ludivine. 2018. *Discourse Markers and (Dis)fluency: Forms and Functions across Languages and Registers*. Amsterdam: John Benjamins.
- Crible, Ludivine, Lisbeth Degand and Gaëtanelle Gilquin. 2017. The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast* 17/1: 69–95.
- Drommel, Raimund H. 1980. Towards a subcategorization of speech pauses. In Hans-Werner Dechert and Manfred Raupach eds. *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton, 227–238.
- Dumont, Amandine. 2018. *Fluency and Disfluency. A Corpus Study of Non-native and Native Speakers (Dis)fluency Profiles*. PhD Thesis: Université Catholique de Louvain.
- Foster, Pauline and Parvaneh Tavakoli. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning* 59/4: 866–896.
- Fox, Jon and Sanford Weisberg. 2019. *An R Companion to Applied Regression* (third edition). Thousand Oaks CA: Sage.
- Fruehwald, Josef. 2016. Filled pause choice as a sociolinguistic variable. In Helen Jeoung ed. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from New Ways of Analyzing Variation (NWAV) 22* (second edition). Pennsylvania: Penn Graduate Linguistics Society, 41–49. <https://repository.upenn.edu/handle/20.500.14332/45126> (2 January, 2025.)

- Fuller, Janet M. 2003. Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua* 22: 185–208.
- Fung, Loretta and Ronald Carter. 2007. Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics* 28/3: 410–439.
- Ganzeboom, Harry B.G. 2010. A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007. Paper presented at *Annual Conference of International Social Survey Programme*, Lisbon, May 1 2010.
- Goldman-Eisler, Frieda. 1961. A comparative study of two hesitation phenomena. *Language and Speech* 4/1: 18–26.
- Götz, Sandra. 2013. *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins.
- Grant, Lynn E. 2010. A corpus comparison of the use of *I don't know* by British and New Zealand speakers. *Journal of Pragmatics* 42/8: 2282–2296.
- Greenbaum, Sidney and Gerald Nelson. 1996. The International Corpus of English (ICE) Project. *World Englishes* 15/1: 3–15.
- Greenwell, Brandon M. 2017. pdp: An R package for constructing partial dependence plots. *The R Journal* 9/1: 421–436.
- Gries, Stefan Th. 2020. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory* 16/3: 617–647.
- Gries, Stefan Th. 2021. *Statistics for Linguistics with R* (third edition). Berlin: Mouton de Gruyter.
- Gries, Stefan Th. 2024. Against level-3-only analyses in corpus linguistics. *ICAME Journal* 48/1: 1–25.
- Hasselgren, Angela. 2002. Learner corpora and language testing: Smallwords as markers of learner fluency. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 143–173.
- Kachru, Brach. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In Randolph Quirk and Henry G. Widdowson eds. *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press, 11–30.
- Kirjavainen, Minna, Ludivine Crible and Kate Beeching. 2022. Can filled pauses be represented as linguistic items? Investigating the effect of exposure on the perception and production of *um*. *Language and Speech* 65/2: 263–289.
- Kjellmer, Göran. 2003. Hesitation. In defence of ER and ERM. *English Studies* 84/2: 170–198.
- Laserna, Charlyn M., Yi-Tai Seih and James W. Pennebaker. 2014. *Um ... who like says you know*: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology* 33/3: 328–338.
- Leuckert, Sven and Sofia Rüdiger eds. 2021. *World Englishes 40/4: Special Issue on Discourse Markers and World Englishes*.
- Lounsbury, Floyd G. 1954. Transitional probability, linguistic structure and systems of habit-family hierarchies. In Charles E. Osgood and Thomas A. Sebeok eds. *Psycholinguistics: A Survey of Theory and Research Problems*. Bloomington, IN: Indiana University Press, 93–101.
- Loy, Jia E., Hannah Rohde and Martin Corley. 2017. Effects of disfluency in online interpretation of deception. *Cognitive Science* 41/6: 1434–1456.

- Maclay, Howard and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *WORD: Journal of the International Linguistics Association* 15: 19–44.
- Manhire, Laura. 2021. *Yeah nah she'll be right: An attitudinal study of 'yeah nah' in New Zealand English*. MA Thesis. Canterbury: University of Canterbury dissertation.
- Miller, Jim. 2009. *Like* and other discourse markers. In Pam Peters, Peter Collins and Adam Smith eds. *Comparative Studies in Australian and New Zealand English*. Amsterdam: John Benjamins, 315–336.
- O'Connell, Daniel C. and Sabine Kowal. 2005. *Uh* and *Um* revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research* 34/6: 555–576.
- OECD. 1999. *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*, OECD Publishing, Paris, <http://www.oecd.org/education/1841854.pdf> (23 May, 2024.)
- Osborne, John. 2013. Fluency, complexity and informativeness in native and non-native speech. In Gaëtanelle Gilquin and Sylvie De Cock eds. *Errors and Disfluencies in Spoken Corpora*. Amsterdam: John Benjamins, 140–161.
- Pawley, Andrew and Frances H. Syder. 1983. Two puzzles for linguistic theory: Native-like selection and native-like fluency. In Jack Richards and Richard W. Schmidt eds. *Language and Communication*. London: Longman, 191–226.
- Pons Bordería, Salvador. 2018. The combination of discourse markers in spontaneous conversations: Keys to undo a gordian knot. *Revue Romane* 53/1: 121–158.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Reichelt, Susan. 2021. Recent developments of the pragmatic markers *kind of* and *sort of* in spoken British English. *English Language and Linguistics* 25/3: 563–580.
- Revis, Melanie and Tobias Bernaisch. 2020. The pragmatic nativisation of pauses in Asian Englishes. *World Englishes* 39/1: 135–153.
- Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14/4: 423–441.
- Scheuringer, Andrea, Ramona Wittig and Belinda Pletzer. 2017. Sex differences in verbal fluency: The role of strategies and instructions. *Cognitive Processing* 18/4: 407–417.
- Schow, Ronald L., John M. Christensen, John M. Hutchinson and Michael Nerbonne. 1978. *Communication Disorders of the Aged: A Guide for Health Professionals*. Baltimore, MD: University Park Press.
- Schweinberger, Martin. 2018. The discourse particle *eh* in New Zealand English. *Australian Journal of Linguistics* 38/3: 395–420.
- Segalowitz, Norman. 2010. *Cognitive Bases of Second-Language Fluency*. London: Routledge.
- Shriberg, Elizabeth. 1996. Disfluencies in switchboard. *Proceedings of the International Conference on Spoken Language Processing (ICSLP '96)*, Volume addendum, 11–14. Philadelphia, PA, 3–6 October. <http://www.asel.udel.edu/icslp/cdrom/vol3/1031/a1031.pdf> (2 January, 2025.)
- Sokołowski, Andrzej, Ernest Tyburski, Anna Sołtys and Ewa Karabanowicz. 2020. Sex differences in verbal fluency among young adults. *Advances in Cognitive Psychology* 16/2: 92–102.
- Strobl, Carolin, James D. Malley and Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and

- regression trees, bagging and random forests. *Psychological Methods* 14/4: 323–348.
- Stubbe, Maria and Janet Holmes. 1995. *You know, eh* and other ‘exasperating expressions’: An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication* 15/1: 63–88.
- Svartvik, Jan. 1990. The TESS project. In Jan Svartvik ed. *The London Lund Corpus of Spoken English: Description and Research*. Amsterdam: Rodopi, 203–218.
- Tagliamonte, Sali. 2005. *So who? Like how? Just what?* Discourse markers in the conversations of young Canadians. *Journal of Pragmatics* 37/11: 1896–1915.
- Tottie, Gunnel. 2011. *Uh* and *Um* as sociolinguistic markers in British English. *International Journal of Corpus Linguistics* 16/2: 173–197.
- Tottie, Gunnel. 2015. *Uh* and *um* in British and American English: Are they words? Evidence from co-occurrence with pauses. In Rena Torres Cacoullos, Nathalie Dion and André Lapierre eds. *Linguistic Variation: Confronting Fact and Theory*. New York: Routledge, 38–55.
- UNESCO. 2006. International Standard Classification of Education (ISCED 1997). Paris: UNESCO. https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf (2 January, 2025.)
- Van Gelderen, Elly. 2010. *An Introduction to the Grammar of English: Revised Edition*. Amsterdam: John Benjamins.
- Vine, Bernadette. 2016. Pragmatic markers at work in New Zealand. In Lucy Pickering, Eric Friginal and Shelley Staples eds. *Talking at Work: Communicating in Professions and Organizations*. London: Palgrave Macmillan, 1–25.
- Weiss, Elisabeth M., Daniel Ragland, Colleen M. Brensinger, Warren B. Bilker, Eberhard A. Deisenhammer and Margarete Delazer. 2006. Sex differences in clustering and switching in verbal fluency tasks. *Journal of the International Neuropsychological Society* 12/4: 502–509.
- Wolk, Christoph, Sandra Götz and Katja Jäschke. 2021. Possibilities and drawbacks of using an online application for semi-automatic corpus analysis to investigate discourse markers and alternative fluency variables. *Corpus Pragmatics* 5/1: 7–36.
- Wright, Marvin N. and Andreas Ziegler. 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77/1: 1–17.
- Wright, Marvin N., Andreas Ziegler and Inke R. König. 2016. Do little interactions get lost in dark random forests? *BMC Bioinformatics* 17: article 145. <https://doi.org/10.1186/s12859-016-0995-8> (2 January, 2025.)

Corresponding author

Sandra Götz
 Philipps University Marburg
 Department of English and American Studies
 Wilhelm-Röpke-Str. 6
 35032 Marburg
 e-mail: goetz-lehmann@uni-marburg.de

received: November 2024
 accepted: January 2025