

English Language and Linguistics, 25.3: 563–580. © The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
doi:[10.1017/S1360674321000253](https://doi.org/10.1017/S1360674321000253)

Recent developments of the pragmatic markers *kind of* and *sort of* in spoken British English¹

SUSAN REICHELT

University of Konstanz

(Received 13 October 2020; revised 27 May 2021)

This study reports on recent changes in the use of the hedges *kind of* and *sort of* in spoken British English over the past twenty years. A quantitative analysis of these features within subsets of the original BNC 1994 (BNC Consortium 2007) and BNC 2014 (Love *et al.* 2017) suggests a systematic encroaching of *kind of* into contexts that are traditionally occupied by *sort of*. This is highlighted in apparent-time patterns in which younger speakers are leading in use as well as real-time patterns that show a significant increase in use between 1994 and 2014.

The hedges *sort of* and *kind of* are often treated as semantically equivalent, yet show distributional differences across different varieties of English. This article reports on an ongoing shift in the use of *kind of* as well as a relatively stable use of *sort of*. Its main focus is a detailed sociolinguistic analysis of both variants, which, in addition to social factors involved, teases apart some of the linguistic aspects of this shift.

In line with the theme of this special issue, the article draws attention to the usefulness of comparable, or comparably made, corpora that allow for focused studies of linguistic change across speakers, generations, registers and communities.

Keywords: hedge, pragmatic marker, diachronic, synchronic, age

1 Introduction

This study investigates recent changes in the use of the pragmatic markers *kind of* and *sort of* across different age cohorts, genders and linguistic contexts in spoken British English. In previous linguistic research, pragmatic markers have been described as fuzzy items (Lakoff 1975: 234), as regards both meaning and function. The linguistic forms that are included in the broad group of pragmatic markers (e.g. *like*, *well*, *you know*, *I mean*, *so*) are generally defined as carrying ‘little obvious propositional meaning’ (Beeching 2016: 23) and operating predominantly on interactional levels of conversation. The items are syntactically flexibly employed and prone to fast-paced change, not just in terms of frequencies, but also in terms of the varying syntactic contexts in which they appear, their semantic uses and their social meanings.

¹ This study was funded as part of the project ‘*The British National Corpus* (BNC) as a sociolinguistic dataset: Exploring individual and social variation’, ESRC grant no. EP/P001559/1.

The pragmatic markers *kind of* and *sort of* have been studied before, albeit not as extensively as other pragmatic markers, such as *like* (e.g. Buchstaller 2006; D'Arcy 2007, 2017; Buchstaller & D'Arcy 2009) or *you know* (Erman 2001; Macaulay 2002; Fox Tree & Schrock 2002). *Kind of* and *sort of* as pragmatic markers are, for the purpose of this article, conceptualised as markers that add hedging (i.e. mitigation, downtoning, vagueness) to an utterance or specific parts of an utterance (see section 2 for a summary of functional definitions).

Tracing changes in use and function is challenging for a variety of reasons, from defining the envelope of variation (i.e. the variable context; see Wiltschko *et al.* 2018), to disambiguating the functions, or finding enough markers in comparable contexts to make systematic judgments on their uses. The increasing availability of diachronic corpora of spoken language is a meaningful asset in collecting evidence of ongoing change. This article serves as a first step to looking at the most recent developments of the pragmatic markers *kind of* and *sort of* and their uses as shaped by social and linguistic factors. A second step will be to investigate syntactic factors in more detail, which is included here as a preliminary analysis (see section 4).

In order to describe its use and development, *kind of* is investigated alongside the 'near synonymous' *sort of* (Gries & David 2007: 1; see also Aijmer 1984: 118; Mauranen 2004: 179), which is the preferred variant in British English varieties (Miskovic-Lukovic 2009: 619). Research has found that there might be syntactic preferences determining the variation between these two hedges (cf. Gries & David 2007), further detailed in section 4. Miskovic-Lukovic (2009: 619) provides an in-depth analysis interpreting when the hedges *kind of* and/or *sort of* are used and, importantly for the current study, whether the inclusion of both *kind of* and *sort of* under one umbrella of variation is indeed warranted. While noting some possible semantic differences, she claims that, structurally and functionally, these items are used similarly, and they seem to have developed in comparable ways (Miskovic-Lukovic 2009: 607–8). The present study thus investigates the development of *kind of* in direct comparison to its nearest equivalent, also to see whether a change in one of them might affect the other. The study aims to address two research questions:

1. How has the use of the pragmatic markers *kind of* and *sort of* changed over the course of twenty years?
2. Which factors correlate with the observed changes?

A brief account of the two markers, their functions and uses is followed by an introduction of the data and methods used in the study. The analysis focuses first on frequency distribution across age and gender, both salient factors for linguistic change; this is followed by a brief syntactic analysis, and, finally, a multivariate analysis that statistically confirms the observations of change.

2 Pragmatic markers *kind of* and *sort of*

In some previous studies, *kind of* and *sort of* have been looked at as individual items (Aijmer 2002; Fung & Carter 2007; Beeching 2016), while other studies have covered

both variants alongside each other (Gries & David 2007; Miskovic-Lukovic 2009). Yet other studies have focused on particular aspects of their uses, such as their role within the construction of general extenders, e.g. *and that sort of thing* (Aijmer 2002; Cheshire 2007). Similar to the contested definition of the whole group of pragmatic markers, the exact categorisation and functional definition of these two variants is still unclear (but see below for an attempt at a functional summary). For instance, it might be argued that one of the key defining features of pragmatic markers – the possibility of omissibility from discourse without changing the propositional content (see example (1)) – is not always fulfilled. As hedges, *kind of* and *sort of* might indeed affect the meaning of the modified item, as shown in example (2) below.

- (1) but then I just *sort of* think well there's four of us (BNC2014, female, age 19)
 *but then I just think well there's four of us
- (2) I'm *kind of* retired now and I'm j just like a show dog you know just like er b but we're *kind of* old now and just like settled down (BNC2014, female, age 29)
 *I'm retired now and I'm j just like a show dog you know just like er b but we're old now and just like settled down

In (2), *kind of* is used as a hedge to modify the two adjectives *retired* and *old* to diminish the impact these two words might have on the listener and the speaker's attitude towards their progressing age. By downtoning the following word, the meaning it carries in unmodified form is adapted and, as can be argued, modified in the process. Further functions are discussed in the remainder of this section.

Their apparent link to interpersonal meaning-making, politeness and face-saving strategies, mitigation and hedging has established *kind of* and *sort of* as typical pragmatic markers. Like many other markers, they carry very little propositional content but have retained semantic links to the lexical roots from which they have derived. Aijmer concludes that 'the hedging meaning of "sort of" is therefore seen as an extension or change from the more literal meaning of "a sort of (a type of)"' (2002: 180), a process also applicable to *kind of*. Both forms have developed from binominal constructions with subcategorisation meaning (Denison 2002), in which the type noun (*sort/kind*) functions as the head of the phrase and is followed by *of* and a second noun. Over time, these constructions developed postdeterminer constructions, where 'the string *sort/kind of* together with the primary determiner forms a complex determiner' (Brems & Davidse 2010: 181). Here, the head of the binominal construction (*kind/sort*) is reanalysed and forms a determiner unit (primary determiner followed by *kind of/sort of*) that precedes the second noun, and as a result, the head-status of *kind/sort* is lost. Examples of these forms from this study's corpus (see section 3 for more information about the corpus) can be seen in (3) and (4).

- (3) people who like that *kind of* music (BNC 1994, male, age 41)
 (4) that's the *sort of* problem (BNC1994, male, age 51)

I call this the 'type construction' use of *sort of/kind of*, and consider its function as 'propositional'. Brems & Davidse (2010: 91) summarise different constructions with

type nouns and point to the development of ‘pragmatic’ functions, such as hedging, since type nouns already indicate ‘that the description is only approximate’. Through the process of grammaticalisation (Hopper & Traugott 1993), the propositional meaning of subcategorisation expands to include qualifying constructions, which carry pragmatic functions such as mitigation (Brems & Davidse 2010: 181).

Miskovic-Lukovic (2009) summarises the development of *kind of* and *sort of* and notes that descriptions of the forms have covered shifts from noun constructions to adjectives (*kind/sort of* as adjectival modifiers of nouns), adverbs (*kind/sort of* as adverbial modifiers of verbs, adjectives, adverbs) and finally pragmatic particles (2009: 609). This is indicated initially by the loss of the plural morpheme for *sort/kind* and, at a later point, the loss of the article preceding the construction. Following this development, *sort/kind + of* appeared as qualifying modifiers in contexts other than noun phrases (see example (6)). Further, both variants also include phonologically reduced items in spoken English: *kinda* and *sorta* (see Brinton 1996: 33–5).

In the following, using examples from the present study’s corpus, I illustrate known pragmatic or discourse functions of *kind of* and *sort of*, first introduced by Beeching (2016: 192–3). This categorisation (i) to (iv) is used as the basis for the present study. We see that the corpus results reflect the variants’ functional repertoire and are thus representative of the range of potential functions that they can serve within the context of spoken British English.

i. Metacommenting, hedging and qualifying

Metacommenting, according to Aijmer (2002: 209), refers to linguistic distancing ‘from the responsibility for using words which are inappropriate because they are technical, trite, too informal, too formal, etc.’. The functions of hedging and qualifying relate to the previously discussed binominal uses where the modified construction is merely an approximation of the actual meaning of the modified item.

(5) er you know you’re mixing with you know highly sophisticated you know *kind of* cultured people in in in big project usually but you know step outside and meet more people from the street you know (BNC2014, female, age 38)

(6) it’s Friday today so everyone’s *sort of* relaxed and good (BNC1994, male, age 14)

In example (5), the speaker seems aware that what she is saying might be taken the wrong way, so by adding *kind of* to the adjective ‘cultured’ she implies that the phrase is not to be taken too seriously. In (6), *sort of* approximates the meaning of ‘relaxed’ to be understood as ‘for a workday, this is not as stressful as a Monday’. The speaker is hedging ‘relaxed’ so as not to appear too cavalier at work.

ii. Mitigating face threats

This function relates to processes of politeness and saving one’s own or somebody else’s face in discourse. In example (7) below, the speaker has inserted *sort of* in order to mitigate a possible insult, that is, the inference that the interlocutor, without the new and expensive

glasses, looks unintelligent. He immediately follows up with another *sort of*, possibly to find a safer description for the other person.

(7) they make you make you look *sort of* intelligent *sort of* (BNC1994, male, age 44)

iii. Pause-filling

Pause-filling, in my data, includes examples of self-repair and general discourse particles that are not intended as fillers, but rather as communicative facilitators.

(8) give us a catch and stuff like that but then that was *kind of* er you know that was *kind of* it it wasn't really (BNC2014, female, age 27)

In (8), the speaker appears to be at a momentary loss for words, also noticeable through the hesitation marker *er*, repetition, self-repair and the pragmatic marker *well*.

iv. General extender

The final function mentioned by Beeching (2016: 193) is that of general extenders, which insert the variants *sort of* and *kind of* into a relatively stable construction: *and/or + sort/kind of + thing/stuff*. She points out that this function overlaps with face-threat mitigators; such overlaps highlight the difficulty in precisely defining functions of these variants.

(9) A: and what did you like doing? (BNC2014, female, age 19)

B: erm like badminton tennis (BNC2014, female, age 19)

A: yeah (BNC2014, female, age 19)

B: running just like long distance hiking *kind of* thing? (BNC2014, female, age 19)

In (9), the speaker seems to be saying that 'long distance hiking' might not be quite the right category for the type of sport she participates in.

Further to Beeching's (2016) functions, I have also found the variants to be used in quotative function as in examples (10) and (11).

(10) and he *sort of* oh what can I get you today mate? (BNC2014, male, age 66)

(11) the next night her car was nicked from the carpark but she didn't seem all she she was *sort of* oh yeah someone nicked my car but she didn't seem indi[?] no I think she'd arranged to have that nicked (BNC2014, female, age 49)

The use of *sort of* and *kind of* in this function is relatively sparse, with only 11 tokens across over 8,500 relevant pragmatic marker variants. The approximation indexed by the marker signals less reliable reported speech and indicates to the listener that what is being retold as speech is only a near-accurate report.

Both *kind of* and *sort of* serve interpersonal and politeness functions in the sense that the propositional content is less strongly asserted and fuzzier compared to the same utterance without these forms. Beeching (2016: 191–3) notes that distinguishing among these functions is nearly impossible without detailed manual coding and prosodic analysis. I would also question whether it is possible to attempt to neatly categorise individual tokens, as they might indeed serve more than one function in the

same instance. For the present study, I focus on the general pragmatic function of *kind of* and *sort of* rather than a smaller-scale approach that would include a more detailed functional investigation. The overall goal is to obtain an initial analysis of ongoing change of these forms using a large corpus, before analysing the data on a functional level in future studies.

3 Methodology

3.1 Data

The data used in this study come from two subsets created from the spoken sections of (a) *The British National Corpus* (BNC) from 1994 and (b) the BNC 2014. The resulting corpus is nearly equally balanced across different categories (gender, region, socioeconomic background, word count contribution) and specifically designed to support research into linguistic change. Speakers who are included in the subset corpus range from age 5 to 95 with years of birth spanning a century (1899 (BNC1994) and 2009 (BNC2014)). This allows for a combination of apparent-time and real-time approaches, which disambiguates possible interpretations of generational change and age-grading, which is often difficult with studies following only one approach (see below).

The study aims at obtaining as full a picture as possible of the use of *kind of* and *sort of* across all age cohorts to allow for different approaches and comparability with future studies. Thus, in order to put the subset corpus together, some of the meta-information from the original corpora had to be re-coded; most importantly, age groups were aligned to create seven age cohorts (see table 1). This differs from the parent corpora, which do not use similar age delineations, and from the original BNC, which does not distinguish age differences beyond the age of 60.

Linguistic change is often found first with typical innovators: usually (female) adolescents around the age of 17 (Stenström 1999, 2002; Chambers 2003; Tagliamonte & D'Arcy 2009; Tagliamonte 2016). Older speakers, in comparison, exhibit relative stability in their stylistic repertoire (Labov 2001; Wagner 2012); this stability has led to a lack of linguistic interest and, as Coupland notes, is a reflection of 'ageist stereotypes of older adults being "set in their ways"' (2001: 191–3). Elsewhere, he warns that a main focus on younger speakers associates older age cohorts with 'an unmarked demographic condition' (2004: 69) and prevents diverse and balanced speaker representation in linguistic research (see also Barbieri 2008: 59; Pichler *et al.* 2018: 2). Further, by extending the age ranges, apparent-time and real-time approaches can be combined to allow for a more thorough investigation. The apparent-time approach involves the study of linguistic change at a single point in time instead of comparing language use as time passes, i.e. the real-time approach. The assumption behind the apparent-time approach is that different age cohorts reflect the language of their respective generations and can be compared. Thus, higher frequencies in use of a particular feature by younger speakers might indicate a change in progress, i.e. a new feature coming into use. However, age-based variation is not a sure sign of ongoing

Table 1. *Subset corpus distribution across gender and age*

Social variables		BNC1994	BNC2014	Total
Gender	Female	126	126	252
	Male	124	124	248
Age	0–14	16	12	28
	15–24	47	56	103
	25–34	50	60	110
	35–44	47	31	78
	45–59	46	50	96
	60–74	31	31	62
	75–95	13	10	23

change, as Wagner (2012: 371) highlights; apparent-time studies might overestimate variation phenomena such as age-grading as generational change. Real-time studies look at language change across different points in time. If this approach includes only particular age cohorts, e.g. only adolescents, the results might lead to misinterpretation of the data. Thus, combining approaches and including a wide glance across the age scale allows for a better analysis, ‘with the relative strengths of one approach offsetting the weaknesses of the other’ (Bailey 2002: 330). With this in mind, the subset corpus aims at including age cohorts at representative and comparable levels, although it must be noted that balancing efforts relied heavily on the distribution already existing in the parent corpora.

The second main speaker variable that is included in this study is gender. Due to lack of further information about the original BNC, it must be assumed that speakers contributing data to the corpora all identified as either male or female. While in the case of the BNC 2014 speakers were asked to indicate their gender on a free-text form, allowing for non-binary, a-gender, or other gender identities to be named, all participants identified a binary identity as either male or female (Robbie Love, personal correspondence). Table 1 presents a summary of the numbers of speakers in the age cohorts and genders in the two subcorpora.

Both subcorpora consist of 250 speakers, with 2,714,337 words from the original BNC and 5,938,032 words from the BNC 2014. Only speakers with at least 1,000 contributed words and complete meta-data were added to the subsets. The differing word counts in the subcorpora were addressed in the analyses by providing relative frequencies (in percentages and occurrences per 1,000 words).

Other social information embedded in the subset corpus is regional background (nine regions across the British Isles), occupation and socioeconomic status. However, the present study focuses on age and gender only.

3.2 Procedure

All tokens for *kind of*, *kinda*, *sort of* and *sorta* were extracted using the POS-tagged version of the subset corpus in Sketch Engine (Kilgarriff *et al.* 2014), including the

immediate context (100 characters) to the left and right to allow for initial disambiguation. For the analysis, the phonologically reduced forms were included with their respective base forms. As previously mentioned, the distinction between the propositional and pragmatic functions of *kind of* and *sort of* in noun phrases is marked syntactically by the position of the determiner. Because of possible expansion into new positions, however, all occurrences required semi-manual coding despite seemingly clear part-of-speech patterns. All tokens were coded for whether they were used as a pragmatic marker (including all functions as discussed in section 2) or as a propositional type noun construct. A second round of coding was then applied to all pragmatic markers, where it was determined what part of the sentence the marker modified.

Overall, 13,706 tokens were extracted, of which 5,164 were propositional and 8,541 were pragmatic uses. The following analysis provides a general overview of the distribution and overall development of the pragmatic markers, followed by a more detailed look into age-based variation indicating a possible linguistic change.

4 Results and discussion

4.1 General distribution

A total of 8,541 pragmatic marker tokens were included in the analysis, of which 3,079 are *kind of* and 5,462 are *sort of*. This reflects previous findings that suggested that *sort of* is the preferred variant in British English varieties (cf. Miskovic-Lukovic 2009; Beeching 2016). Both variants increased in relative use between 1994 and 2014, although the level of increase shows a stark contrast and points towards a noticeable uptake of *kind of*. Some more details on this development can be gained from the discourse values of both forms, which reflect the markers' pragmatic function potential. Discourse values (d-values) are relative representations of a marker's 'discourse function in relation to grammatical function expressed (in percent)' (Stenström 1990: 161; Aijmer 2002: 27). Based on the claim that the propositional use is a rather stable value, a shift in discourse values thus indicates a shift in the pragmatic use of a form (Beeching 2016: 77), pointing towards a feature undergoing change. The discourse value, represented in percentages, serves mainly as a normalisation method that enables comparison across different studies and different markers at various points in their development. For the present study, the above-mentioned uptake of *kind of* as a pragmatic marker is clearly visible (see table 2).

In 1994, the discourse value for *kind of* is still relatively low (24 per cent), with 76 per cent of occurrences of all *kind of* tokens of the type noun constructions, but the discourse uses rise to 63 per cent in 2014. The discourse value for *sort of* is relatively stable in comparison (60 per cent in 1994 and 56 per cent in 2014); whether the slight decrease in pragmatic functions correlates to the increase of *kind of* will be shown in the more detailed age brackets in section 4.2.

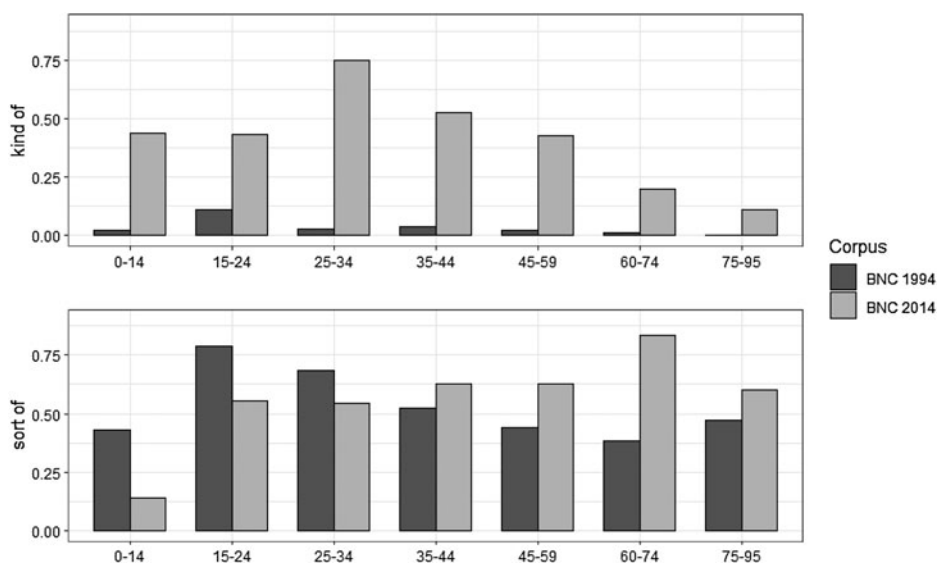
Table 2. *Distribution of kind of and sort of in the subset corpus*

	1994			2014		
	N	per 1,000 words	d-value	N	per 1,000 words	d-value
<i>kind of</i>	59	0.04	24%	3,020	0.48	63%
<i>sort of</i>	1,415	0.55	60%	4,047	0.6	56%

4.2 Social factors

In the following, the apparent increase of *kind of* in relation to *sort of* will be observed across different age groups and at two points in time. Data here are represented as relative per 1,000 words.

Figure 1 shows the mean distribution of *kind of* across the seven age cohorts (upper panel). Data from 1994 reveal very low use with a slight increase in the age cohort 15 to 24, which fits neatly into the linguistic innovator bracket. Twenty years later, the lighter grey bars reveal that the use of *kind of* by this group (now aged 35 to 44) has increased even more (from 0.11 to 0.53). This observation holds true for all other generations as well, clearly indicating a change in progress on the level of generational increase as well as lifespan change (Sankoff 2004). This means that the change is not just happening generation by generation, but that the use of *kind of* is increasing within each generation as well.

Figure 1. *Kind of* and *sort of* across age, per 1,000 words

In 2014, the most frequent use of pragmatic marker *kind of* appears in the age group 25 to 34, which is slightly beyond the adolescent innovation peak, but still within the range of young speakers. The slight dip in the youngest speaker groups (0–14 and 15–24) might give an indication of added social meaning that the considerable increase has caused. If that is the case, it means that the variant has developed into what Labov has defined as a marker, i.e. forms ‘that have attracted sufficient attention to emerge ... in stylistic variation’ (Eckert 2008: 463). *Kind of* might thus be indexically tied to the generation preceding the younger speakers, motivating the latter to avoid using it as a form of stylistic distinction. However, to speak with any kind of certainty of such an interpretation, attitudinal data would be needed.

Looking at the age cohorts for *sort of* (lower panel), we can detect a relatively stable use across the generations. The highest use in 1994 appears, again, with speakers aged 15 to 24. By 2014, however, this has shifted to the age group 60 to 74, which is well beyond expected innovative ages. Further, for all younger age cohorts (0 up to 34), the use of *sort of* actually decreased between 1994 and 2014. Still, despite the noticeable increase of *kind of*, *sort of* remains the overall preferred variant and is relatively stable and unaffected. This is at odds with previous studies claiming that both variants are serving similar functions, as such a drastic increase in one variant would have replaced the other (the division of labour principle). This might then point to a structural expansion, i.e. where the increase of *kind of* also points to a new application.

Figure 2 shows the distribution across year of birth, simulating a view of both corpora neatly overlapping rather than being separated by twenty years. It indicates user-specific frequencies of the variants (y-axis) alongside a density count, which indicates where most speakers are located in terms of their uses. *Kind of* (upper panel) shows the earliest use as a pragmatic marker (in this dataset) by speakers born in the 1920s (in comparison, *sort of* is used by one of the oldest speakers included in the corpus, born in 1899). Here, the development of the form becomes quite clear, with the highest uses of pragmatic marker *kind of* appearing among speakers born between 1970 and 1990.

The lower panels of *sort of* show that the form is used more often altogether, but that the high point of use based on the year of birth lies somewhere between the 1960s to 1980s. Its relative stability is further shown by few outliers and an almost indistinguishable shift between the two datasets.

The juxtaposition of change and stability of variants that are said to be functionally equivalent is further highlighted in a second observation of discourse values (see table 3). As above, the values represent the relative use of pragmatic forms against all forms (pragmatic + propositional). Unsurprisingly, the discourse value increased within each age cohort for *kind of*, supporting the idea that this feature has gained new uses, either in pragmatic functions or syntactic contexts. The higher discourse value also indicates higher levels of societal acceptance. Beeching (2016: 194), when commenting on consistently high discourse values of *sort of* in her study, says it shows ‘that pragmatic marker *sort of* is a highly respectable way of hedging speech’. Returning to the interpretation of slightly lower uses of *kind of* in the youngest two age cohorts of 2014, another possibility might be that it is precisely the gained acceptability indicated through

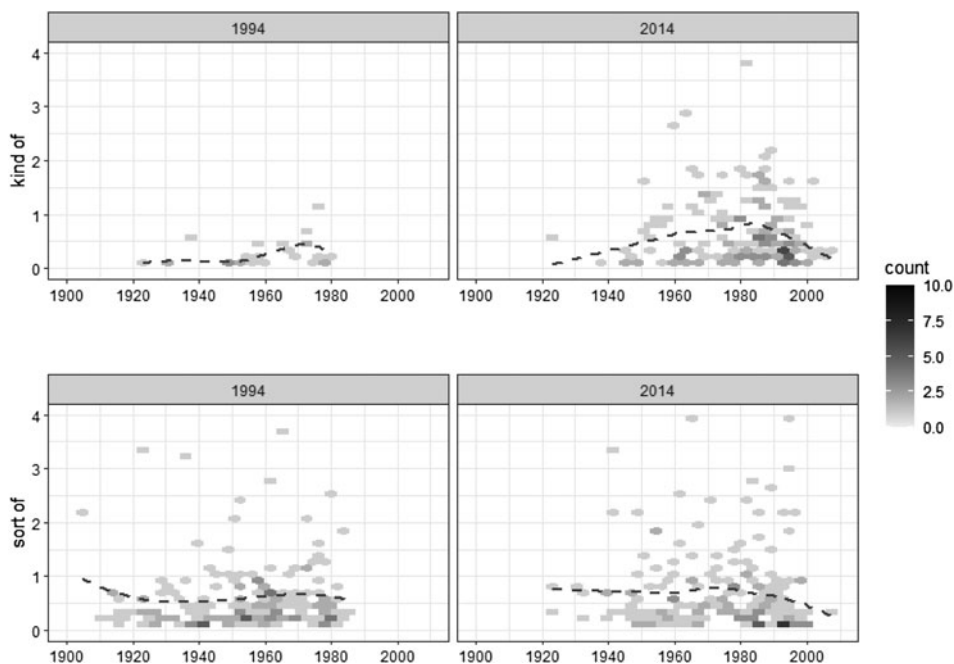


Figure 2. *Kind of* and *sort of* across year of birth, per 1,000 words

Table 3. *Discourse value of kind of and sort of across age groups (percentage)*

Age group	<i>kind of</i>		<i>sort of</i>	
	1994	2014	1994	2014
0–14	33	84	59	63
15–24	34	68	50	66
25–34	22	64	59	58
35–44	29	53	63	62
45–59	21	60	56	59
60–74	18	52	49	56
75–95	0	75	55	55

the high discourse value (84 and 68 per cent) that prevents further high uses in that age bracket. As with other pragmatic markers (e.g. intensifiers), certain functions demand a swift changeover to remain meaningful in discourse (cf. Tagliamonte 2008).

Sort of exhibits only one notable increase in discourse values, which is located in the age group 15 to 24. This distribution is quite likely to be an indicator of age-graded structural behaviour rather than any linguistic change.

The fact that change in progress is usually spearheaded not only by adolescents, but by female adolescents, warrants a brief investigation of the distribution across gender. In

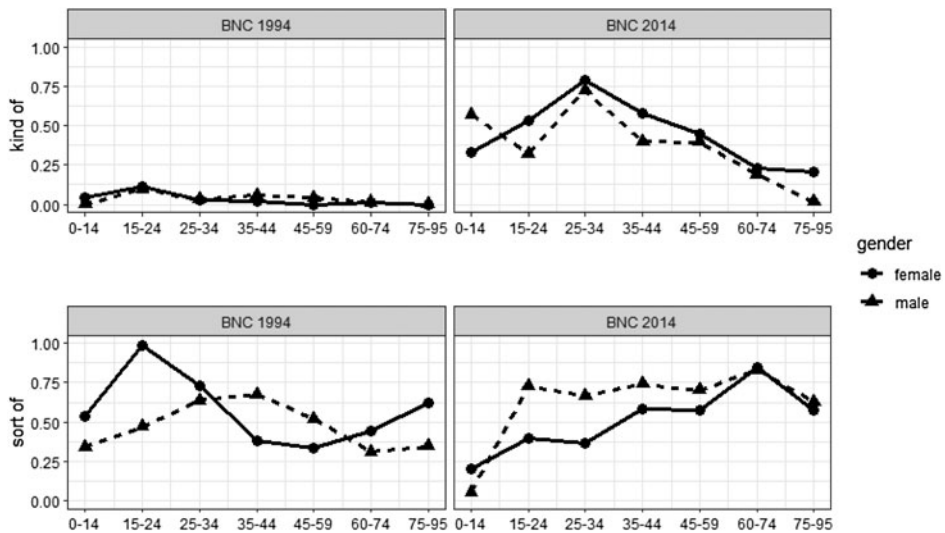


Figure 3. *Kind of* and *sort of* across age and gender, per 1,000 words

figure 3, female speakers are indicated with a solid line, while male speakers are indicated with a dotted line. In 1994, we can see that for the youngest three age cohorts, female speakers are slightly leading in use of *kind of*, which, for the following cohorts, is reversed. By 2014, female speakers are more consistently leading in use, particularly in age groups 15 to 24 and 35 to 44. For the youngest speakers, a different pattern appears, which may be explained by my previous interpretation of social meaning attached to incoming variants. The distribution does not show any statistically significant gender differences ($p > .05$).

In comparison, the distribution for *sort of* (lower panel) appears much more unstructured. This seems likely for a feature that is not particularly marked and widely accepted across age and gender categories. It appears that male speakers in 1994 exhibit an expected stratification where the highest use can be found in the younger age cohorts (15 up to 44) before tapering off with increasing age. By 2014, this pattern has widened further to include ages up to 74, indicating that the pragmatic marker *sort of* is no longer age marked at all. Even more interesting is the distribution for female speakers, which almost inverts between 1994 and 2014. In 1994 we find a distribution rather typical for age-grading, with high uses throughout adolescence and a quick decrease with increasing age. By 2014, however, this pattern has shifted completely with the highest uses now recorded in the age groups 60 to 74 and 75 to 95. It seems as though, in 2014, young women have found alternative forms of hedging (certainly *kind of*, as indicated above, but possibly also alternative variants). Across all age groups and both corpus subsets, their use of *sort of* is lowest of all. This distribution is particularly remarkable considering that it is this age bracket that has shown the highest discourse value, possibly pointing to a decrease in the use of propositional *sort of*.

Taking the rapid change of *kind of* in frequencies into account, as well as possible shifts in use as indicated by discourse values, the next step is to look at the immediate context of the pragmatic markers.

4.3 *Internal factors*

As previously mentioned, this part of the analysis will be kept relatively short, as a more detailed and fine-grained exploration of precise functions would call for prosodic analyses. This part of the analysis thus provides a first investigation of the type of modification found with *kind of* (*kinda*) and *sort of* (*sorta*), followed by a summarising multivariate analysis.

Most of the coding for contextual information was enabled by the POS-tagged format of the subset corpus, which allowed me to categorise tokens in a straightforward manner. The initial step was to evaluate tokens that appeared with a determiner immediately to the left of the node as in (12), which almost always indicate propositional uses and were thus excluded from the study. Slightly more complex alternatives of propositional uses include determiner + adjective + node, as in (13), and were also excluded from the pragmatic focus.

(12) the *kind of* books I'm reading (BNC2014, female, age 20)

(13) it might erm be the right *kind of* book (BNC1994, female, age 33)

Pragmatic uses that were included in the study are modifications of adjectives (14), adverbs (15), nouns (16), prepositions (17) and verbs (18).

(14) it's just *kind of* warmish (BNC1994, female, age 34)

(15) he just walks *kind of* calmly (BNC2014, male, age 21)

(16) somebody who was selling *kind of* kits of that (BNC2014, female, age 64)

(17) she flew up in the air *sort of* across the kitchen (BNC2014, female, age 49)

(18) we *kind of* made profit (BNC1994, female, age 44)

Further, pragmatic markers are also included when they appeared clause-finally (19), or were used in other pragmatic functions, such as approximation (20), quotation (21), or as a slightly more ambiguous discourse particle (22), which includes filler functions, and self-repair items.

(19) he's gone home *sort of* (BNC1994, male, age 14)

(20) I managed to get on a bus at *sort of* five-ish (BNC1994, female, age 28)

(21) he was *sort of* oh god (BNC2014, male, age 66)

(22) Well he just totally *sort of* because I think he knows (BNC1994, female, age 34)

Thus, the markers are quite varied in syntactic context and the question arises whether any of these contexts have shifted with the increased use of *kind of*. Gries & David (2007: 9) found that, when looking at collocations with parts of speech, *kind of* showed preference in the modification of nouns and adjectives, while *sort of* appeared more often than expected with adverbs and verbs, as well as whole propositions. Additionally, they

Table 4. *Syntactic application of kind of and sort of, relative value of modified instances (percentage)*

	<i>kind of</i>		<i>sort of</i>	
	1994	2014	1994	2014
Adjective	22.9	12.1	8.1	9.1
Adverb	0.0	4.7	4.0	4.5
Approximator	0.0	0.8	4.7	3.4
Clause-final	2.1	0.7	1.3	1.1
Noun	16.7	17.0	13.0	15.1
Preposition	10.4	13.7	9.5	11.9
Quotative	0.0	0.2	0.2	0.1
Verb	47.9	50.9	59.2	54.8

note that, in terms of modified lemmas, the two variants share little collocational overlap. This means that the variants are following ‘lexically determined variation’ (2007: 11), whereby *kind of*, for instance, more likely modifies adjectives describing emotional states or verbs describing mental activities. This pattern might explain how *kind of* increased so drastically while *sort of* remained relatively stable. *Kind of* is expanding within the system; however, it is not doing so by taking over contexts that are traditionally modified by *sort of* as both variants seem to follow distinct modification patterns.

Table 4 summarises the modification for both variants at both points in time. Because of ambiguity in function, the variants that were coded as discourse particles (see example (22)) are not included. In 1994 *kind of* modified four categories of parts of speech (nouns, prepositions, verbs and adjectives) and was used in clause-final position. By 2014, it had gained uses in all remaining syntactic contexts as well, which explains the increased discourse values. The contexts of *sort of* appear largely unchanged. In terms of distinct uses, the two variants do not appear to be dividing their uses across modifiable and other pragmatic contexts. That means that both variants are used as modifiers across different parts of speech, albeit with different preferences; see discussion of multivariate analysis below for a statistical evaluation of expected and realised preferences.

Finally, in order to test the factors involved in the analysis, a multivariate regression analysis in Rbrul (Johnson 2009) was run to determine significance levels and overall trends in preference between the use of *kind of* and *sort of*. As previously established, both forms can be considered equivalent in function with slightly differing use preferences. As pragmatic markers, the two variants form an envelope of variation, or variable context, which allows for a variationist approach. The following analysis tests which constraints, social and linguistic, are statistically meaningful when *kind of* is chosen over *sort of*.

Table 5. *Multivariate analysis of kind of and sort of*

<i>kind of vs sort of</i>				
Input prob.	0.428			
Total N	8541			
Deviance	6,386,338			
	log-odds	Factor weight	%	N
Age		p<3.38e-11		
	-0.0578			8,541
Corpus		p<1.31e-39		
2014	2.132	0.894	43	7,067
1994	-2.132	0.106	4	1,474
Gender		Not significant		
Female	0.178	0.544	43	4,883
Male	-0.178	0.456	27	3,658
Modification		p<3.40e-07		
Clause-final	0.398	0.598	27	75
Adjective	0.351	0.587	44	777
Noun	0.268	0.567	40	1,188
Preposition	0.234	0.558	41	933
Adverb	0.156	0.539	38	346
Discourse particle	0.0506	0.513	34	844
Quotative	0.0222	0.506	46	13
Verb	-0.0258	0.494	34	4,160
Approximator	-1.454	0.189	10	205
Speaker as random factor				
	Intercept	N	Proportion	
Std.dev	2.205	8,541	0.36	

Factors included reflect the previous distributional analysis: age as a continuous variable, gender, syntactic contexts and of course the respective subset corpus. Speakers were included as random factors, which accounts for individual preferences that might otherwise indicate an unbalanced sample (see [table 5](#)).

The analysis shows that the most significant constraint is the point in time, i.e. the subset corpus, which confirms that the general change is indeed an ongoing generational change in progress. The data show that *kind of* is more likely to be used in the 2014 corpus. This confirms a general change in the use of *kind of*. Age is the second most significant predictor for the use of *kind of* over *sort of*. That is, with decreasing age, the chances of a speaker choosing *kind of* as a pragmatic marker are increasing. Syntactic context is also found to be significant in constraining the occurrence of *kind of*. Similar to the findings by Gries & David (2007), *kind of* favours modification of nouns and adjectives, while disfavouring verbs. Finally, gender was not found to be significant, which reflects the findings discussed previously. Men and women, at both points in time as well as across age cohorts, roughly follow the same pattern, with women leading only slightly in frequency. This conclusion echoes

Beeching's (2016: 209) study of *sort of* in British English contexts, where she discusses several social patterns, including gender, and finds no difference in frequency or function between male and female speakers. Nonetheless, the gendered use of *sort of* in 2014 (see figure 3) indicates a pattern that will be interesting to observe in more detail.

5 Conclusion

The analysis shows that the pragmatic marker *kind of* has increased considerably over the span of just twenty years, from 1994 to 2014. The approach of combining apparent-time and real-time studies further highlights that age-based uses of the marker correlate with patterns of age-grading, generational change, as well as lifespan change. *Sort of*, as a marker serving similar functions, has previously been established as the preferred British English variant. While this preference is still upheld in the 2014 subset corpus, the data suggest that the two variants, in British English contexts, are no longer distinguished. Surprisingly, general frequencies of *sort of* have not been affected as might be expected (although it appears as though the marker is losing its prominence with younger speakers). The recent increase of *kind of* also raises questions of possible indexed social meaning, as became clear in the visualisations of development across time and gender.

In terms of the linguistic structures that the variants modify, the markers appear to have slight preferences for certain parts of speech. A cautionary glance into the future might include structural separation and more pronounced linguistic differentiation. A more detailed follow-up study of functional and structural preferences might untangle further patterns of pragmatic language change.

The study not only sheds light on the development of a pragmatic marker variable in spoken British English but also highlights the usefulness of comparable larger-sized corpora that provide full speaker meta-data (detailed information on chronological age alongside year of birth). By being able to put together subsets of material to allow for the type of approach that fits the research questions, scholars can trace language change more thoroughly. Second and third-generation corpora, such as the two BNC sets, enable linguists to investigate language change in more detail and with more rigour.

Author's address:

Advanced Data and Information Literacy Track

Fachgruppe Medienwissenschaft

University of Konstanz

Postfach 157

Universitätsstr. 10

78457 Konstanz

Germany

susan.reichelt@uni-konstanz.de

References

- Aijmer, Karin. 1984. 'Sort of' and 'kind of' in English conversation. *Studia Linguistica* 38(2), 118–28.
- Aijmer, Karin. 2002. *English discourse particles: Evidence from a corpus*. Amsterdam and Philadelphia: John Benjamins.
- BNC Consortium. 2007. *The British National Corpus*, version 3 BNC XML edition. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. www.natcorp.ox.ac.uk
- Bailey, Guy. 2002. Real and apparent time. In J. K. Chambers, Peter Trudgill & Natalie Schilling (eds.), *The handbook of language variation and change*, 312–32. Malden, MA: Blackwell.
- Barbieri, Federica. 2008. Patterns of age-based linguistic variation in American English. *Journal of Sociolinguistics* 12(1), 58–88.
- Beeching, Kate. 2016. *Pragmatic markers in British English: Meaning in social interaction*. Cambridge: Cambridge University Press.
- Brems, Lieselotte & Kristin Davidse. 2010. The grammaticalisation of nominal type noun constructions with *kind/sort of*: Chronology and paths of change. *English Studies* 91(2), 180–202.
- Brinton, Laurel J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.
- Buchstaller, Isabelle. 2006. Social stereotypes, personality traits and regional perception displaced: Attitudes towards the 'new' quotative in the UK. *Journal of Sociolinguistics* 10(3), 362–81.
- Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multilocal, multivariate investigation of quotative *be like*. *Journal of Sociolinguistics* 13(1), 291–331.
- Chambers, J. K. 2003. *Sociolinguistic theory: Linguistic variation and its social significance*, 2nd edn. Malden, MA: Blackwell.
- Cheshire, Jenny. 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11(2), 155–93.
- Coupland, Nikolas. 2001. Age in social and sociolinguistic theory. In Nikolas Coupland, Srikant Sarangi & Christopher N. Candlin (eds.), *Sociolinguistics and social theory*, 185–211. London: Pearson.
- Coupland, Nikolas. 2004. Age in social and sociolinguistic theory. In Jon F. Nussbaum & Justine Coupland (eds.), *Handbook of communication and aging research*, 69–90. New York: Routledge.
- Denison, David. 2002. History of the *sort of* construction family. Presented at the Second International Conference on Construction Grammar (ICCGS), University of Helsinki, Finland. www.humanities.manchester.ac.uk/medialibrary/llc/files/david-denison/Helsinki_ICCG2.pdf (accessed 20 May 2021).
- D'Arcy, Alexandra. 2007. *Like* and language ideology: Disentangling fact from fiction. *American Speech* 82(4), 386–419.
- D'Arcy, Alexandra. 2017. *Discourse-pragmatic variation in context: Eight hundred years of LIKE*. Amsterdam and Philadelphia: John Benjamins.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4), 453–76.
- Erman, Britt. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 33(9), 1337–59.
- Fox Tree, Jean E. & Josef C. Schrock. 2002. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34(6), 727–47.
- Fung, Loretta & Ronald Carter. 2007. Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics* 28(3), 410–39.

- Gries, Stefan Th. & Caroline David. 2007. This is *kind of* / *sort of* interesting: Variation in hedging in English. In Päivi Pahta, Irma Taavitsainen, Terttu Nevalainen & Jukka Tyrkkö (eds.), *Studies in variation, contacts and change in English*, vol. 2: *Towards multimedia in corpus studies*. www.helsinki.fi/varieng/journal/volumes/02/gries_david/
- Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1), 359–83.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vit Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1), 7–36.
- Labov, William. 2001. *Principles of language change: Social factors*. Malden, MA: Blackwell.
- Lakoff, George. 1975. Hedges: A study in meaning criteria and the logic of fuzzy concepts. In D. J. Hockney, William Harper & B. Freed (eds.), *Contemporary research in philosophical logic and linguistic semantics*, 221–71. Dordrecht: Springer.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3), 319–44.
- Macaulay, Ronald. 2002. You know, it depends. *Journal of Pragmatics* 34(6), 749–67.
- Mauranen, Anna. 2004. ‘They’re a little bit different.’ Observations on hedges in academic talk. In Karin Aijmer & Anna Stenström (eds.), *Discourse patterns in spoken and written corpora*, 173–97. Amsterdam and Philadelphia: John Benjamins.
- Miskovic-Lukovic, Mirjana. 2009. *Is there a chance that I might kinda sort of take you out to dinner?* The role of the pragmatic particles *kind of* and *sort of* in utterance interpretation. *Journal of Pragmatics* 41(3), 602–25.
- Pichler, Heike, Suzanne Evans Wagner & Ashley Hesson. 2018. Old-age language variation and change: Confronting variationist ageism. *Language and Linguistics Compass* 12(6), e12281.
- Sankoff, Gillian. 2004. Adolescents, young adults and the critical period: Two case studies from ‘Seven Up’. In Carmen Fought (ed.), *Sociolinguistic variation: Critical reflections*, 121–39. Oxford: Oxford University Press.
- Stenström, Anna-Britta. 1990. Pauses in monologue and dialogue. In Jan Svartvik (ed.), *The London–Lund Corpus of Spoken English: Description and research*, 211–52. Lund: Lund University Press.
- Stenström, Anna-Britta. 1999. *He was really gormless – she’s pretty crap*: Girls, boys, and intensifiers. In Hilde Hasselgård & Signe Oksefjell (eds.), *Out of corpora: Studies in honour of Stig Johansson*, 69–78. Amsterdam and Atlanta, GA: Rodopi.
- Stenström, Anna-Britta. 2002. *Trends in teenage talk: Corpus compilation, analysis and findings*. Amsterdam and Philadelphia: John Benjamins.
- Tagliamonte, Sali A. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12(2), 361–94.
- Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. & Alexandra D’Arcy. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85(1), 58–108.
- Wagner, Suzanne Evans. 2012. Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6(6), 371–82.
- Wiltschko, Martina, Derek Denis & Alexandra D’Arcy. 2018. Deconstructing variation in pragmatic function: A transdisciplinary case study. *Language in Society* 47(4), 569–99.