

The social embedding of a syntactic alternation: Variable particle placement in Ontario English

MELANIE RÖTHLISBERGER 

University of Zurich

SALI A. TAGLIAMONTE 

University of Toronto

ABSTRACT

The present work investigates the effects of social constraints on word order variation in particle placement in Ontario English, Canada. While previous research has documented numerous linguistic factors conditioning the choice of variant, social correlates have so far remained unexplored. To address this gap, we analyze 6,047 variable phrasal verbs from the vernacular speech of six communities in Ontario. These data were coded for length of the direct object, verb semantics, community, and the individual's education, gender, age, and occupation. Our analyses confirm previous findings that variation in particle placement is predominantly determined by direct object length. However, we also expose significant social and geographic factors, and importantly an effect of age, with younger speakers using the joined variant more than older speakers. Further analysis confirms that the latter effect is consistent across communities, indicating a change in progress, possibly due to ongoing grammaticalization of particles in the verb phrase.

The present study investigates the effect of contextual factors—linguistic and social—on word order variation in particle placement, as in (1) in Ontario, Canada. While word order alternations such as the present one have been extensively studied in the literature, we know very little about their social embeddedness in well-defined speech communities and the extent to which patterns of variation correlate with speakers' social background. The present study addresses this gap by assessing the effect of social and prominent language-internal constraints on particle placement in one specific language area.

Phrasal verbs are defined here as partially lexicalized verb-particle combinations with the particle being of adverbial nature (Rodríguez-Puente, 2017:71; Thim, 2012:10). More specifically, we are interested in transitive phrasal verbs which take a direct object as complement either following the verb-particle combination (as in 1a) or intervening between the verb and the particle (as in 1b) including also compositional phrasal verbs (see Rodríguez-Puente, 2017:72).

- (1) a. *joined order*
 Just **pick**_[verb] **up**_[particle] people_[direct object],
 start throwing them. (Thaddeus Bickley, 30, Kirkland Lake)¹
- b. *split order*
 Oh yeah, yeah, I used to **pick**_[verb] people_[direct object] **up**_[particle].
 (Jason Gill, 37, Toronto)

Semantically, phrasal verbs can be placed along a continuum from semantically opaque forms that constitute a single lexical unit whose meaning is noncompositional, that is, not inferable from its components (see Schneider, 2004:230), such as *give up* in *I'd give up television in a minute*, to purely compositional phrasal verbs such as *take out* as in *They took out the beds* (see also Ishizaki, 2012:241–2).

The two variants in (1) are generally considered semantically equivalent but pragmatically and stylistically different. Such stylistic or pragmatic factors constraining the choice between the joined and the split order have received ample attention in the literature (see, among others, Cappelle, 2006, 2009; Dehé, 2002; Grafmiller & Szmrecsanyi, 2018; Gries, 2003; Haddican & Johnson, 2012; Lohse, Hawkins, & Wasow, 2004; Rodríguez-Puente, 2016, 2017, 2019). In contrast, the effect of speaker-related social factors (e.g., individual's age or gender) on this choice has so far been ignored (but see Kroch and Small [1978], which investigates the effect of prescriptivism). The present study thus goes beyond traditional analyses by considering social as well as language-internal constraints on word order variation. With this approach, we follow a recent upsurge of studies that have demonstrated that syntactic alternations are contextually constrained both by factors that are inherent to the linguistic system and those that arise out of language usage and which are socially motivated (see Geeraerts, Kristiansen, & Peirsman, 2010:7–8; also Röthlisberger, 2020).

The paper is structured as follows: The next section sketches the historical trajectory of phrasal verbs from Old English to present-day English. After that, we provide details on the sampled data and the extraction process as well as the statistical methodologies employed. Next, we present the results and discuss their implications. The last section offers final conclusions and directions for future research.

PHRASAL VERBS: PAST AND PRESENT

The historical trajectory

Phrasal verbs have undergone substantial syntactic and semantic development from the Old English (OE) period to present-day English (Claridge, 2000:83; Rodríguez-Puente, 2017:69). In OE, verb-particle combinations allowed for both pre- and postverbal position of the particle without any semantic differentiation (see Claridge, 2000:84–5). In both positions, intervening elements between the verb and the particle were permissible for transitive phrasal verbs (see Claridge [2000:84] citing Hiltunen [1983]). Semantically, phrasal verbs seem to have

expressed some kind of motion with the particle adding a directional or spatial meaning to the construction while metaphorical meanings can also be found (e.g., *forþfēran* ‘to travel/move away or by’ but also figuratively ‘to die’) (Thim, 2012:5).

Preverbal positions mostly gave way to postverbal patterning in late OE and early Middle English (Hiltunen, 1983:106–11) likely due to increasing restrictions on word order and the loss of object-verb order (Thim, 2012:103). By the end of the Middle English period, the postverbal position of the particle had become the norm (Claridge, 2000:85). According to Thim (2012:87–8), the development of phrasal verbs between OE and Middle English reflects an ongoing grammaticalization process, namely decategorialization and loss of syntactic freedom and an increase of constructional types with noncompositional meaning. Besides these language-internal changes, language-external influences have also been held accountable for this development. A possible (indirect) influence of Old Norse—which has advanced phrasal constructions—is discussed in Hiltunen (1983), Denison (1985:49–53), and Lutz (1997). Other sources of influence include translation of Latin compound verbs to native English verb-adverb combinations (Claridge, 2000:88), and borrowing of verbs from Norman French (Claridge, 2000:116).

During the Early Modern English period (c. 1500–1800), phrasal verbs increased in productivity (Brinton, 1988:187), particularly phrasal verbs with noncompositional, that is, idiomatic, meaning (Claridge, 2000:96). This increase proceeded in an interrupted rather than a continuous manner; declines in usage have been ascribed to the influence of prescriptivism, especially during the eighteenth century (see Claridge, 2000:96–8, 178–9; Rodríguez-Puente, 2019:175). From the nineteenth century onward, both the number of different phrasal verbs (tokens) and the number of unique verb-particle combinations (types) increase in frequency again (Thim, 2006:218). Speech-related text types (such as fiction) were particularly hospitable to these developments (Brown & Palmer, 2015:80).

With regard to alternating phrasal verbs specifically, a diachronic perspective is offered by Elenbaas (2013) and Rodríguez-Puente (2016). Elenbaas’ analysis shows that in Middle English the joined order is preferred with nominal direct objects and the split order with pronominal direct objects. Overall, the joined order is more frequent than the split order (Elenbaas, 2013:493) and this trend continues into Early Modern English (Elenbaas, 2013:495) and into the Late Modern English period up until present-day English, now even with pronominal objects (Rodríguez-Puente, 2016:150). In contrast to these studies, Gries (2003) observed that the split order predominates in spoken British English at the end of the twentieth century, while the joined order is preferred in written language (see also Cappelle, 2006:9).

The synchronic perspective

Research studying alternating phrasal verbs in contemporary English has exposed multiple contextual constraints influencing this variation, including type of object (pronoun versus NP) and length of the direct object (see, e.g., Grafmiller &

Szmrecsanyi, 2018; Gries, 2003). Indeed, it is important to note that, when the direct object is an unstressed pronoun, the split order is categorical (**pick up it* versus *pick it up*). In contrast, the influence of length is a tendency: as the length of the direct object increases, the joined order becomes more likely until the direct object is sufficiently long to make the split order impossible (Cappelle, 2009). Other contextual constraints that have been considered include discourse-familiarity, definiteness, and persistence of the object. Besides such constraints referring to characteristics of the direct object, the semantics (i.e., idiomaticity or compositionality) of verb-particle combinations have also been shown to be influential, with idiomatic verb-particle combinations, for example, *kick up* ('cause trouble') or *give off* ('emit'), exhibiting a higher preference for the joined variant (e.g., Szmrecsanyi, 2005:132).

Common to the majority of these studies is their limited focus on one speech setting while cross-lectal comparisons (across styles or registers, across dialect regions) have remained rare. Exceptions include Cappelle (2006) and Gries (2003) who highlighted the existence of cross-register differences in that the split variant is significantly more frequent in spoken than in written discourse.² This difference has been ascribed to the rather informal nature of the split order variant (see Bolinger, 1971:57, fn. 8) but it is potentially also due to the split order occurring frequently with pronominal direct objects which are prevalent in spoken language.

Regional variation has only recently received attention. Comparing UK and US twitter data, Haddican and Johnson (2012) observed that UK tweeters prefer the split order significantly more than US tweeters (similar results were observed by Cappelle [2009:165–6] when comparing data in Lohse, Hawkins, & Wasow [2004]). Using data from the Freiburg English Dialect Corpus (FRED) (Hernández, 2006), Szmrecsanyi (2005) observed regional differences in the preference of variant within the UK. Taking a macroperspective on dialectal variation, Schneider (2004) compared patterns of particle placement across five varieties of English, using data from the respective International Corpus of English (ICE) component. He found that second language speakers prefer the joined variant more than native British speakers (Schneider, 2004:239). These tendencies are confirmed in Szmrecsanyi, Grafmiller, Heller, and Röthlisberger (2016) and Grafmiller and Szmrecsanyi (2018).

This recent interest in regional variation notwithstanding, other social factors besides region have remained under the radar. The absence of such predictors is mostly due to the lack of vernacular speech from speakers for which the individuals' social background such as age or gender is known.

DATA AND METHODOLOGY

The corpus

The data for this study come from a multidialectal corpus of conversational interviews from six communities in Ontario, all from the Ontario Dialects Project housed in the Language Variation and Change Research Laboratory at

the University of Toronto. Due to the fact that this corpus samples the common vernacular of everyday speech with large amounts of data (typically an hour or more of conversation) from the same individuals, we can explore the social aspects of variation in particle placement more fully. The six speech communities under scrutiny were selected because they offer the most substantive community-based data sets in the project and have a relatively balanced distribution by year of birth and gender. The main community-level contrast in these materials is the difference between Toronto, the largest urban center of the province, and moderately large towns at varying distance in the outlying areas, as indicated in the map in [Figure 1](#).

The data from Toronto were collected between 2003 and 2006. Data collection in the five northern communities spans 2009 to 2011. The corpus data from these six communities amount to a total of 5,444,278 words.

Extracting the tokens

We used a PERL script to extract phrasal verbs from the raw text files based on a list of ten frequent particles (*around, away, back, down, in, off, on, out, over, up*) (Grafmiller & Szmrecsanyi, 2018:389; also Gries, 2003:203–10) and an extensive list of verb lemmas that have been shown to take part in the variation in a corpus study of nine national varieties of English and 13.5 million words of text (Grafmiller & Szmrecsanyi, 2018).

To find phrasal verbs in the joined order, verbs and particles had to be in immediate proximity; to find the split order, we allowed for six intervening words between the verb and the particle, as this has been shown to be the maximum number of words with which the split order is still possible (Grafmiller & Szmrecsanyi, 2018:389). In a first step, we manually discarded all tokens that followed the surface structure as required by the PERL script but that were not a phrasal verb, for example, *back in* as in *We were waiting at the back in the alley*. Because our search string was fairly open regarding split variants, we also extracted phrasal verbs in the joined order where the verb did not occur in the dataset by Grafmiller and Szmrecsanyi (2018). These were nevertheless retained (full verb list given in [Appendix A](#)).

Defining the envelope of variation

Following closely the methodology in Grafmiller and Szmrecsanyi (2018:389), we next discarded all nonvariable tokens to only retain variable ones, that is, phrasal verbs in either the split or the joined order where the alternating variant was semantically equivalent and grammatically possible. On these grounds, we discarded as nonvariable: intransitive phrasal verbs that were not followed by one nominal direct object (i.e., a noun phrase). This includes all tokens with pronominal direct objects, tokens with two objects, and clausal objects (i.e., finite and nonfinite clauses) (e.g., *everybody found out I was black*). We further excluded verb-preposition combinations where the preposition is part of a prepositional phrase rather than part of the verb phrase

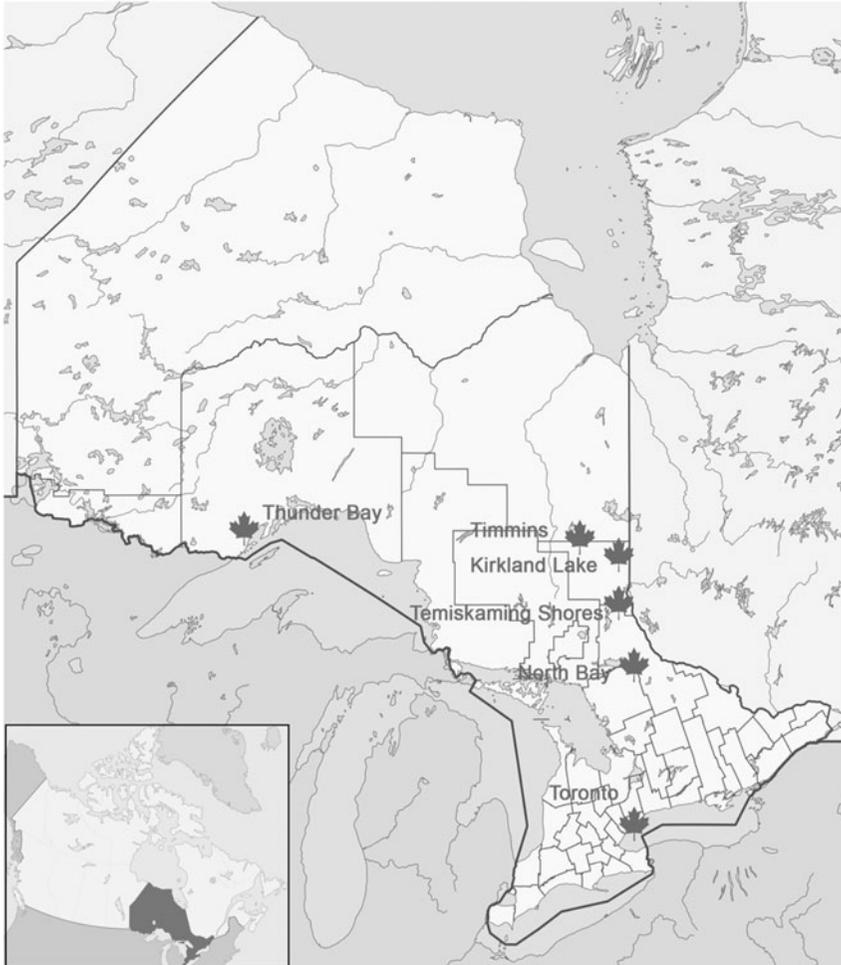


FIGURE 1. Map of Ontario with the six speech communities from which the data were drawn (adapted from Wikimedia, user NordNordWest, original file name: Canada Ontario location map 2.svg, published under a CC-BY-SA 3.0 Unported license, changes made relate to the six communities).

(e.g., *you were driving down a tunnel*), passivized tokens (e.g., *they got wiped out*), phrasal verbs where the direct object was a *wh*-form or a relative pronoun (e.g., *You see the apartment that they put those girls in?*), and tokens with an intervening modifying adverb or two particles instead of one (e.g., *they cleaned the house right up*). Additionally, we excluded prepositional verbs as in (2) since these are not permissible in the split order (see Grafmiller & Szmrecsanyi, 2018:389).

- (2) And he said, “One of them was picking on Jamie.” (Amelia Hannock, 36, Kirkland Lake)

Finally, the dataset was restricted based on length of the longest direct object in the split order (six words) to exclude any phrasal verbs in the joined order with exceedingly long direct objects, as these would make the split variant nearly impossible.

Circumscribing the variable context in this way provided 6,029 potentially alternating variants. These were first manually coded for the direct object and then semiautomatically coded for the factors outlined next.

Annotation of contextual factors

Due to our interest in social factors, we coded only the most prominent language-internal constraints, that is, length of the direct object and idiomaticity of the phrasal verb, in order to focus on social factors, that is, speaker’s year of birth/age, gender, education, occupation, and community. While it could be argued that other language-internal predictors are more important than idiomaticity of the verb, the historical overview above has shown that phrasal verbs with idiomatic meaning increased in frequency over time. It might thus be at the intersection of age (as apparent time construct) and idiomaticity where we can tap linguistic change in progress.

Age

In order to account for the different compilation times of the corpus data, we used year of birth as the most appropriate measure for age. [Figure 2](#) shows the distribution of the joined variants by year of birth of the individual. Due to lack of representation of individuals for every year, we plotted the smoothed conditional means for year of birth to visualize this development over time (see raw counts in [Appendix B](#)). It is immediately apparent that the joined variant increases in tandem with the individual’s year of birth across the twentieth century.

Gender

Gender was coded as a binary predictor distinguishing (perceived) male and female individuals. The proportional distributions shown in [Figure 3](#) indicate that females have a marginally higher proportion of joined variant than males. These differences, even though minimal, are statistically significant ($X^2(1) = 6.30$, $p = .0121$).

Education

As is typically the case in community-based fieldwork, some individuals did not divulge information on their education ($n = 200$). 2,676 tokens came from individuals without secondary education (“N”) and 3,153 tokens from

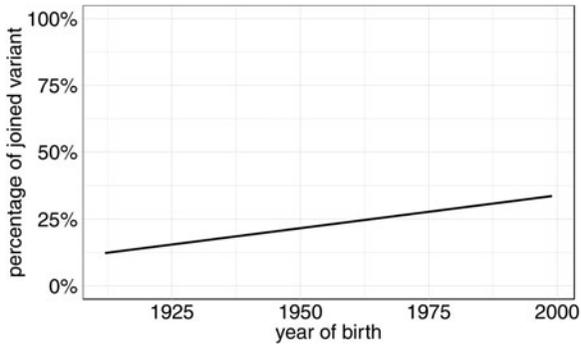


FIGURE 2. Smoothed conditional means of year of birth by percentage of joined variant ($n = 6029$).

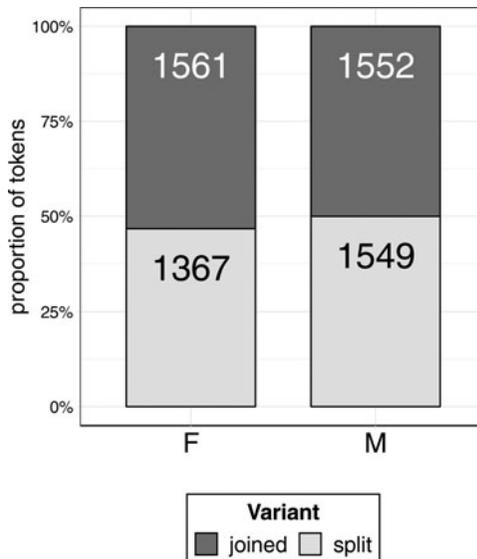


FIGURE 3. Proportional distribution of joined and split variant by gender (F=female, M= male); raw numbers provided in each bar ($n = 6029$).

individuals with secondary education (“Y”). Differences between educational statuses (see Figure 4) are not statistically significant ($X^2(1) = 0.240$, $p = .6242$).

Occupation

Regarding occupation, we distinguished between speakers that have a blue-collar job (“B”), white-collar job (“W”), or students (“S”). As with education, some individuals did not provide information on their occupation

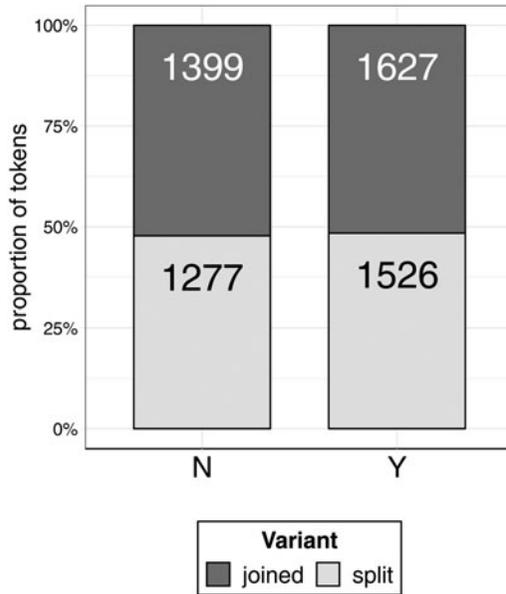


FIGURE 4. Proportional distribution of joined and split variant by education (N = no secondary education, Y = with secondary education; $n = 5829$, NAs not shown).

($n = 476$). The proportional distributions in Figure 5 illustrate that students and white-collar workers exhibit a higher proportion of joined variants than blue-collar workers. We should note, however, that the higher proportion of joined variants in students might be due to (their young) age. Differences are statistically significant between the occupational groups ($X^2(2) = 107.26$, $p < .001$).

Community

The last social predictor, community, was coded on the basis where the speaker was born and/or raised, distinguishing by the six communities sampled in the data (see Figure 1). The proportion of split variants as shown in Figure 6 is slightly higher in Temiskaming Shores and Kirkland Lake compared to the other communities. Differences between communities are statistically significant ($X^2(5) = 105.46$, $p < .001$).

Length of direct object

Length is regularly reported to be the most important predictor in word order variation and is also the most prominent one in particle placement globally (see Grafmiller & Szmecsanyi, 2018:397). We coded length of the direct object in the number of characters instead of words, as this provided a more normal distribution making the data more suitable for statistical analyses (see Appendix C for raw

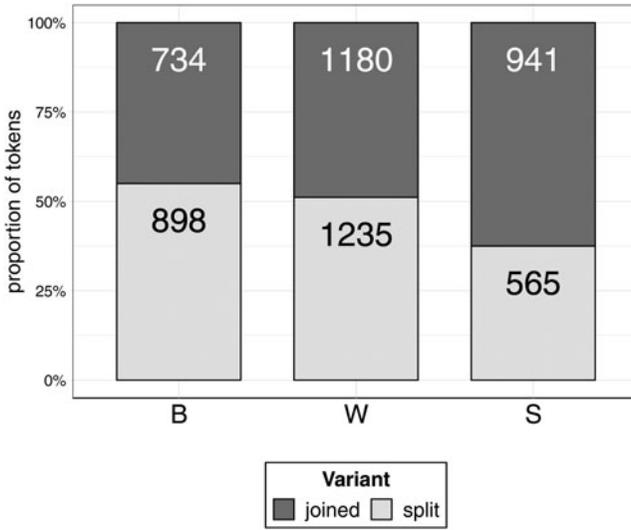


FIGURE 5. Proportional distribution of joined and split variant by occupation (B = blue-collar worker, W = white-collar worker, S = student; $n = 5553$, NAs not shown).

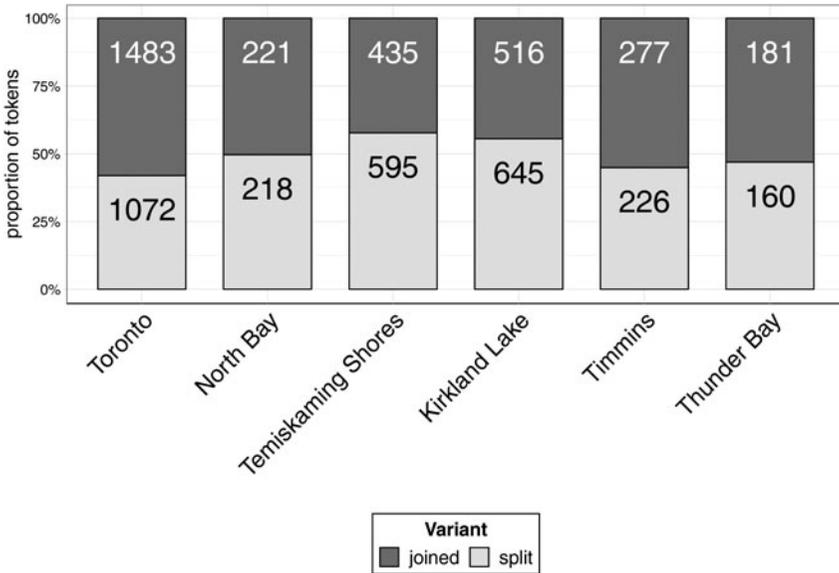


FIGURE 6. Proportional distribution of joined and split variant by place ($n = 6029$).

counts). Figure 7 shows that the proportion of joined variants increases as the direct object increases in length from left to right, which is consonant with earlier work on particle placement (e.g., Grafmiller & Szmrecsanyi, 2018).

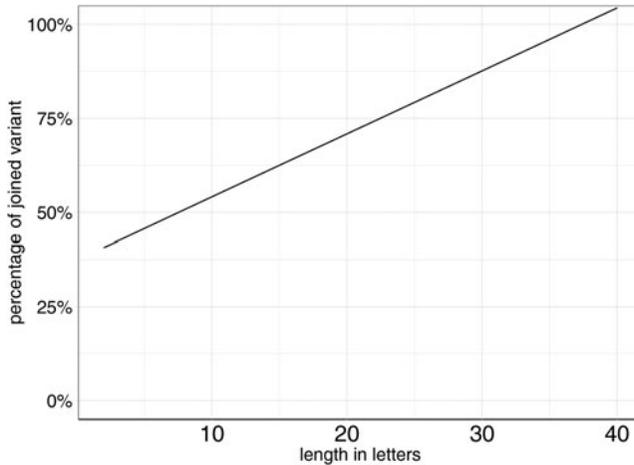


FIGURE 7. Smoothed conditional means of the length of the direct object in the number of characters by percentage of joined variant ($n = 6029$).

Idiomaticity or verb semantics

As pointed out in the historical overview, idiomatic uses of phrasal verbs have been increasing in frequency over time. In order to verify whether such an increase can be substantiated with the current data, we coded for idiomaticity of the phrasal verb. We distinguished between compositional meanings of phrasal verb constructions where the meaning is entirely predictable from that of their parts versus noncompositional or idiomatic uses where the meaning of the token cannot be derived from the separate verb or particle. Following Grafmiller and Szmrecsanyi (2018:392), we made use of the heuristic proposed in Lohse et al. (2004:244–6): if the verb-particle construction [X V (P) NP (P)] entails both [X V NP] and [NP be/become/come/go/stay P], it was considered “compositional”; if the construction failed either of the two tests, it was considered “noncompositional”. Figure 8 shows that noncompositional uses are proportionally more frequent with the joined variant than the split variant, which is plausible since joined variants are more inclined to grammaticalize and thus acquire idiomatic meaning.

Since our focus was on social rather than linguistic constraints, the next step is to explore the variable patterns across different social contexts, examining community-specific developments across time, lexical effects, and, finally, the interplay between social and linguistic factors influencing the choice of variant when all these factors are considered simultaneously.

RESULTS

Community-specific developments across time

Following Szmrecsanyi (2005), who exposed regional differences in his UK data, we took a closer look at community-specific developments in Ontario. To visualize

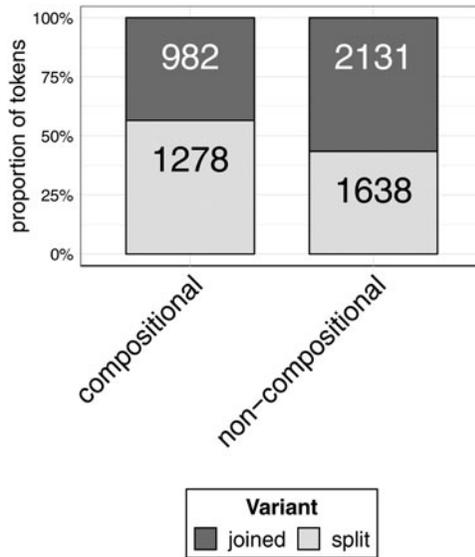


FIGURE 8. Proportional distribution of joined and split variant by idiomaticity ($n = 6029$).

this development across time, we binned year of birth into three groups arrived at by a conditional inference tree fitted on the complete dataset with year of birth as the sole predictor of the variation, using the *partykit* package in R (Hothorn & Zeileis, 2015). The tree split the data as follows: individuals born before 1937, between 1938–1974, and after 1975 (analysis not shown here). Percentages of the joined variant per speech community and per time period are illustrated in Figure 9 (see raw counts in Appendix D).

Figure 9 shows a distinct trend toward the joined variant in apparent time with an accelerating profile after 1974 in most communities. Toronto has the highest percentage of joined variants in all three time periods compared to the other communities, while three communities lag behind Toronto in this development: North Bay, Temiskaming Shores, and Kirkland Lake. In sum, the different places in Ontario do not seem to be a major determinant of variation in particle placement. Rather, the observed increase of the joined variant takes place in all speech communities, with a visible upswing in the late twentieth century and some lag in the frequency of use in the smaller towns in northern Ontario.

Lexical effects

Even though a thorough analysis of lexical effects in particle placement is beyond the present study (but has been done by, for example, Deshors [2016]), we would like to touch upon some lexical considerations that our dataset offers.

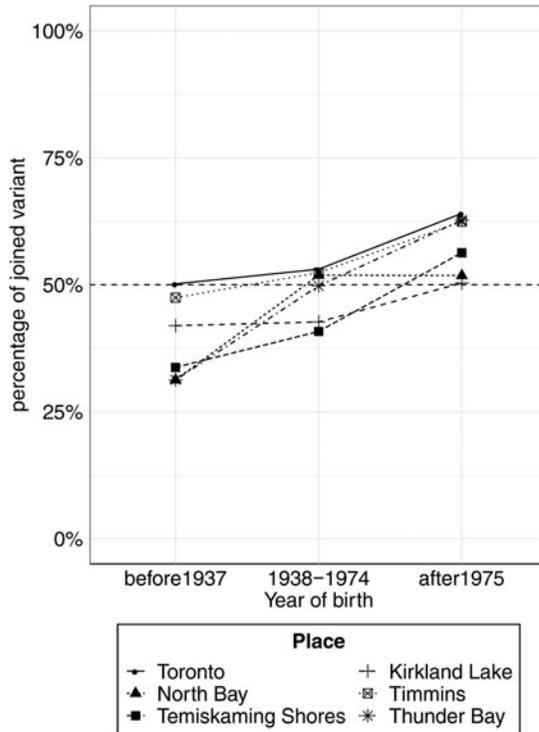


FIGURE 9. Percentage of joined variant by speech community (lines/points) across the three time periods ($n = 6029$).

As pointed out by Brown and Palmer (2015:83) and Rodríguez-Puente (2017:87), certain lexical elements, particles, or verbs have been shown to develop differently over time (see also Ishizaki, 2012). Focusing on the individual lexical elements may thus prove informative when tracing the development of the two variants in order to determine whether the change toward more joined variant might be led by one particular particle or verb. Table 1 shows the ten most frequent phrasal verbs in the corpus with their raw and proportional frequencies by variant. The most frequent phrasal verb in the corpus is *pick up*, followed by *put on*, *put in*, and *take off*. In terms of variants, *pick up*, *put on*, *bring in*, *make up*, *open up*, and *set up* prefer the joined over the split variant while *put in*, *take off*, *take out*, and *put up* prefer the split over the joined variant. A closer look at these ten most frequent verb-particle combinations reveals that they can occur both in noncompositional and compositional uses with extensive variability (see Table 1).

In its idiomatic sense, *pick up*, for instance, can be used for *languages*, *attitudes*, *people*, *speed*, *jobs*, *groceries*, and other items that can be bought. Literal meanings of *pick up* often refer to *the phone*, *the ball*, or *garbage*. In fact, *the phone*

TABLE 1. *The 10 most frequent verb-particle combinations in the data by variant*

verb-particle combination	joined order		split order		Total
	<i>noncompositional</i>	<i>compositional</i>	<i>noncompositional</i>	<i>compositional</i>	
pick up	333 (86.7%)	117	51 (13.3%)	12	384
put on	163 (55.1%)	74	133 (44.9%)	93	296
put in	84 (37.0%)	50	143 (73.0%)	97	227
take off	52 (26.7%)	49	143 (73.3%)	57	195
take out	67 (38.7%)	48	106 (61.3%)	72	173
bring in	94 (62.3%)	52	57 (37.7%)	34	151
put up	66 (44.9%)	26	81 (55.1%)	34	147
make up	111 (86.0%)	0	18 (14.0%)	0	129
open up	100 (87.7%)	0	14 (12.3%)	0	114
set up	105 (92.1%)	2	9 (7.9%)	0	114

constitutes the majority of direct objects used with *pick up*, namely twenty-three times in the corpus (all in the joined order).

At the same time, there seems to be a tendency for joined variants to occur with idiomatic uses, that is, the largest bulk of joined variants of *pick up* are idiomatic, while joined variants of *open up*, *make up*, and *set up* are nearly exclusively idiomatic in usage.

In order to determine whether the increase of joined variants over time is driven by one specific particle, we examined the proportional distribution of each particle by joined and split variant across the three time periods (see the section on “Community-specific developments”). This proportional distribution is shown in Figure 10 with the particles ordered alphabetically from left to right. It is apparent that the joined variant increases over time compared to the split variant for all particles with the exception of *back* and *over*, where the trajectory of change is not linear. Regarding *around*, *down*, *off*, *on*, and *up*, a larger increase is observed from the middle-aged (born between 1938 to 1974) to the youngest speakers (born after 1975) than between the oldest (born before 1937) to the middle-aged speakers, suggesting that these particles might be driving the increase in frequency between 1938 and 1975. *Up* is also among the ten most frequent verb-particle combinations listed in Table 1, especially those verb-particle combinations that prefer the joined over the split variant. For other particles, the increase of the joined variant is similar across the periods (i.e., *away* and *out*) or marginally larger between the oldest and the middle-aged

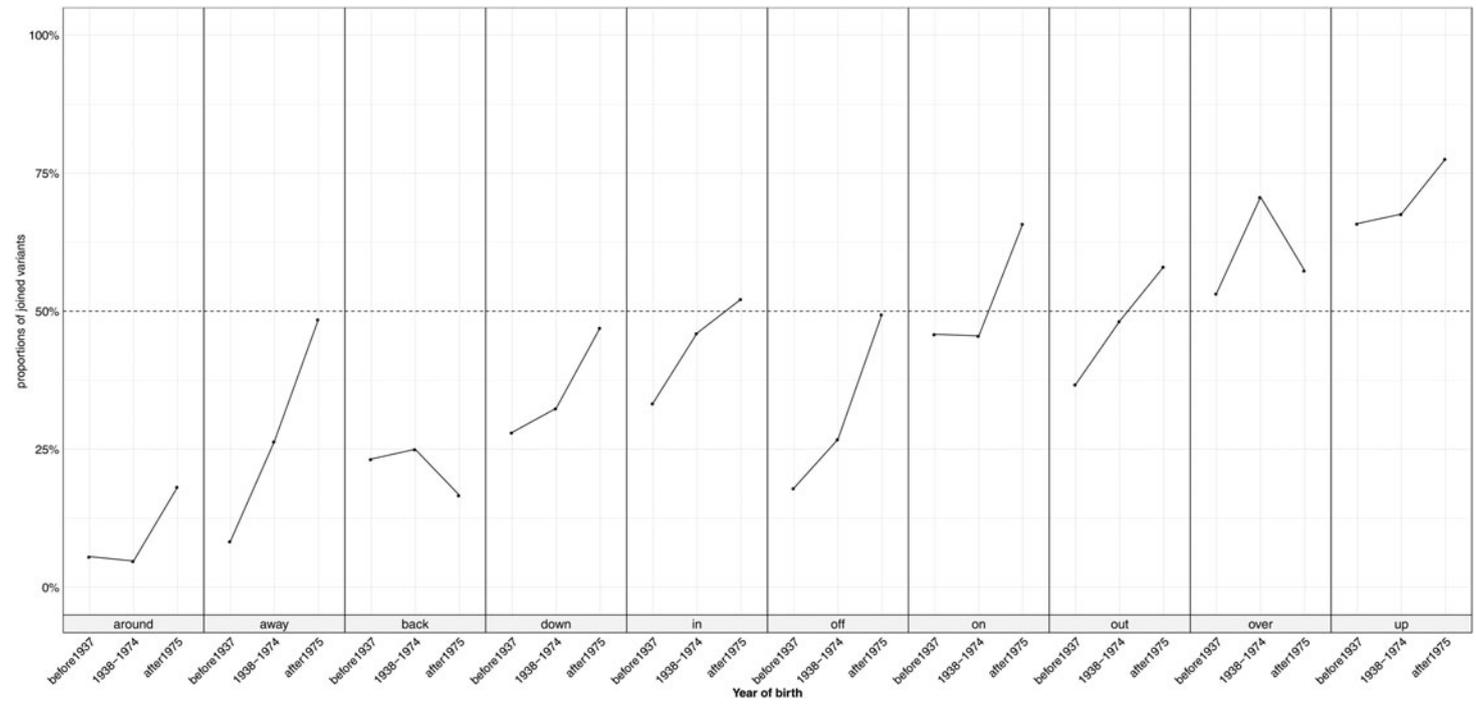


FIGURE 10. Proportions of joined variant by all ten particles per subperiod ($n = 6029$).

speakers (i.e., *in*). In sum, it is not the case that one particle is driving the change toward more joined variant single-handedly; however, *up* seems to be an influential lexical item in that regard.

Effects of social and language-internal constraints

To test the effect of social and linguistic constraints on particle placement, we use mixed-effects logistic regression modeling (Pinheiro & Bates, 2000). Mixed-effects modeling estimates the simultaneous effect of a set of constraints on a binary outcome (here split versus joined variant) and assesses the probability of observing one of the variants based on these constraints. These probabilities, or coefficient estimates, are given on a logit-scale by the underlying mathematical equation (Gelman & Hill, 2007; Hosmer & Lemeshow, 2000). Mixed-effects models take not only the combined set of constraints into account but also allows for so-called random effects: by-group idiosyncratic variation that is specific to the dataset, for example, lexical items, text types, or individuals sampled (Pinheiro & Bates, 2000). Accounting for such idiosyncrasies enables us to better generalize beyond the particular data sample to the population at large. Models were fitted with the *lme4* package in R (Bates, Maechler, & Bolker, 2013; R Core Team, 2017).

To fit the model, we included only phrasal verbs for which all predictors outlined above had information provided, that is, excluding NAs in occupation and education ($n = 5393$). Numeric predictors (i.e., length) were scaled by two standard deviations and centered around the mean, following Gelman (2008). For binary predictors we set the reference level to the level showing a tendency for the split variant according to the proportional distributions shown above (i.e., male, blue-collar worker, etc.) in order to render the model output more interpretable, that is, for all predictors in the model we are predicting the incoming form (i.e., joined variant). Place was coded using deviation contrasts where the proportion of responses for each level is compared against the grand mean across all levels so as not to prejudice the model for any pre-existing theory of geographical diffusion. The three time periods, based on year of birth, were contrast-coded with backward difference coding, which compares each time period against the previous one, rather than the first one (as in treatment coding) (see Menard, 2010:97). This enables us to observe differences from one generation to the next rather than observing differences between the oldest and the youngest speakers.

In order to test for a community effect, we initially also included an interaction of community and age, but the effect did not turn out to be significant and was subsequently left out of the model structure. In order to test whether idiomatic uses of joined phrasal verbs increased over time at the expense of literal meanings and split variants, we also included an interaction between age and idiomaticity, which was significant in the model and thus retained. The random effect structure includes random intercepts for individual and lexical items, that is, particle and verb, grouping verbs occurring less than ten times together to ease model convergence.

TABLE 2. Variance accounted for by random effects in the model

Groups	Variance	Std. Dev
Individual	0.2698	0.5194
Verb lemma	1.0665	1.0327
Particle	0.7704	0.8777

The model performs well on the data: Somer's C index is an excellent 0.85, indicating that the model can discriminate well between the variants (Levshina, 2015:259). The model can predict 76.7% of the data correctly, which is statistically significantly better than the baseline of 51.8%, if predicting randomly the most frequent variant ($p_{\text{binom}} < .001$). Collinearity between the factors in the model was assessed with the condition number κ (following Belsley, Kuh, & Welsch [1980]) which was $\kappa = 7.1$ in the model without the interaction, indicating no collinearity (Baayen, 2008:182).

The random effect structure is shown in Table 2. The largest contribution is made by the verb, followed by the particle and then the individual.

The adjustments of the individual verbs to the intercept, that is, their modeled preference for either variant, is shown in Table 3 with a focus on the six verbs with the highest adjustments to either the joined (positive adjustment) or the split (negative adjustment) variant. Table 3 illustrates that *set*, *pick*, and *fill* prefer the joined variant whereas *get*, *invite*, and *move* strongly prefer the split variant. The preferences of *pick* and *set* for the joined variant fall in line with the results in Table 1, where *pick up* and *set up* have been shown to occur more often in the joined than in the split variant.

Fixed effects are shown in Table 4 with number of instances per level (n), relative frequency of the joined variant, coefficient estimates (β), standard errors (SE), and level of significance (p) for each predictor; estimations are for the joined variant, reference levels are indicated if treatment coding is used. In the case of "Place", each community is compared to the average mean of all communities, which is why no reference level is shown for "Place". Age group always has the previous level as level of comparison, which is why "before 1937" is not the general reference level for this factor.

The results of the model corroborate findings from previous studies in that variation in particle placement is predominantly determined by direct object length. With every unit increase in length (in characters), the odds of a joined variant increase by a factor of 5.4. Confirming the proportional distributions in Figure 2, the model exposes a significant effect of age between speakers born between 1938 and 1974 and those born after 1975: if the speaker is born after 1975, they are two times more likely to use the joined variant than a speaker born between 1938 and 1974; the difference for the oldest age groups is not significant. In terms of place, only Toronto, Temiskaming Shores, and Kirkland Lake deviate from the average across the

TABLE 3. *Adjustments of ten verbs diverging the most from the model’s intercept. Positive adjustments indicate a tendency toward the joined, negative adjustments a tendency toward the split variant*

Verb	adjustment to intercept
set	1.75
pick	1.57
fill	1.56
move	-1.57
invite	-2.67
get	-2.94

TABLE 4. *Mixed-effect logistic regression of the seven constraints tested in the model (n = 5393)*

Factor	Level of categorical factor	n	Rel. freq. joined %	β	SE	p-value
(INTERCEPT)		5393		-0.514	0.330	0.120
L ENGTH		5393		1.689	0.088	<0.001
IDIOMATICITY	compositional	2029	43.8	Reference level		
	noncompositional	3364	56.6	1.120	0.086	0.164
AGE GROUP	before 1937	1038	42.5			
	1938–1974 (compared to “before 1937”)	2203	47.9	0.110	0.183	0.548
	after 1975 (compared to “1938–1974”)	2152	60.2	0.735	0.165	<0.001
PLACE	Toronto	2376	58.2	0.276	0.081	<0.001
	North Bay	349	51.6	0.043	0.152	0.776
	Temiskaming Shores	992	43.0	-0.231	0.106	0.029
	Kirkland Lake	1039	44.5	-0.214	0.107	0.045
	Timmins	319	56.4	0.203	0.155	0.191
	Thunder Bay	328	50.6	-0.077	0.154	0.619
EDUCATION	non	2506	52.6	Reference level		
	yes	2887	51.1	-0.145	0.104	0.164
GENDER	male	2782	50.1	Reference level		
	female	2611	53.6	0.054	0.092	0.555
OCCUPATION	blue-collar	1528	45.5	Reference level		
	white-collar	2359	49.0	0.336	0.125	0.007
	student	1506	62.5	0.408	0.152	0.007
INTERACTIONS						
AGE GROUP : IDIOMATICITY	before 1937 : compositional	413	32.4			
	before 1937 : noncompositional	625	49.1			
	1938–1974 (compared to “before 1937”) : compositional	800	37.4			Reference level for idiomaticity
	1938–1974 (compared to “before 1937”) : noncompositional	1403	54.0	0.041	0.197	0.835
	after 1975 (compared to “1938–1974”) : compositional	816	55.8			Reference level for idiomaticity
	after 1975 (compared to “1938–1974”) : noncompositional	1336	62.9	-0.405	0.158	0.010

whole of Ontario: if a speaker is from Toronto, they are 1.3 times more likely to use the joined variant than the rest of Ontario; if they are from Temiskaming Shores or Kirkland Lake, the odds of a joined variant decreases by a factor of 0.8, that is, speakers from these communities are *less* likely to use the joined variant than the rest of Ontario. Both education and gender were not significant in the model and are hence not discussed in more detail. Idiomaticity does also not significantly impact particle placement (but note that this might be because the factor is included as a higher-order interaction term). Finally, if speakers are white-collar workers instead of blue-collar workers, they are 1.4 times more likely to use a joined variant, and if they are students, they are 1.5 times more likely than a blue-collar worker to use a joined variant. Regarding the interaction between idiomaticity and age, the model indicates that only the last age group (people born after 1975) compared to the middle-aged group show a significant decrease in the use of joined variants with noncompositional meaning (compared to compositional phrasal verbs). However, the difference between the age groups is comparatively small, and idiomaticity as a main effect (in a no-interaction model) does not reach significance. We can thus say that, based on our data, idiomaticity is not significantly constraining particle placement nor does its effect change over all generations.

In sum, the change toward more joined variants over time takes place concurrently in all communities in Ontario, as the interaction testing for community-specific developments was not significant in the initial model. Toronto differs significantly from the rest of Ontario, while the northern communities largely function as a cohesive whole. To further assess whether the effect of occupation is due to an underlying effect of age (i.e., young students), we fitted a second model excluding the student data. The model-fitting followed the same procedure as above with the reduced dataset of $n = 3887$ (results are shown in [Appendix E](#)). Again, the model performs well on the data: Accuracy is 76.7%, which is significantly better than the baseline (52.4% of split variants) ($p_{\text{binom}} < .001$), and Somer's C index is an excellent 0.85. Results of that second model confirm the findings from the first model: young speakers and white-collar workers are more likely to use the joined variant than older speakers and blue-collar workers, and both gender and education are not significant predictors. Interestingly, the interaction between age and idiomaticity is no longer significant. This second model confirms that the change in apparent time observed in the full dataset is not confounded by students, nor by place, and that idiomaticity is indeed only marginally interacting with age.

DISCUSSION

This study aimed to explore the impact of social and geographical constraints on particle placement using a rich, sociolinguistically stratified, community-based

archive of vernacular spoken Ontario English. Analyzing the data with mixed-effects logistic regression for the influence of speakers' age, their occupation, education level, gender, community, idiomaticity, and length of the direct object has shown that the last factor is the most important constraint, consistent with previous findings in the literature (e.g., Grafmiller & Szmrecsanyi, 2018). However, the individual's year of birth, community, and occupation also play a statistically significant role in determining choice of variant.

Importantly, we have documented an ongoing trend toward the joined variant across the twentieth century, which we believe is part of a long line of historical developments consistent with grammatical change, as outlined in the historical overview above. Further analysis of the data confirmed that this trend is parallel across the six Ontario communities and across lexical items. What is more, the statistical modeling has shown that Toronto is ahead of the other communities in terms of preference for the joined variant. This result also corroborates earlier research in Ontario showing that the urban center of Toronto and the largest cities in the north pattern together with respect to linguistic changes, for example, the alternation between pronominal quantifiers in *-body/-one* (Jankowski & Tagliamonte, 2020) and the lexical item *guy* (Franco & Tagliamonte, *in press*). Such geographic trends are well known in historical linguistics, where it has been shown that urban centers are more "liable to language change" (Taeldeman, 2005:276; see also Labov, 2007). Taking all these findings together, we suggest that younger speakers' preference for the joined variant is the synchronic reflex of an ongoing grammatical change that is progressing in parallel across the major varieties of English (e.g., British English, Canadian English). The regularity of this change across varieties, in the absence of a significant gender effect, suggests that the change is largely due to systemic adjustments within the English verb phrase rather than being propelled by social factors. The question is: why is this happening and why are some but not all social factors impacting this variation?

Given the historical record and previous research on this linguistic variable, we might hypothesize the following trajectory in a longitudinal grammaticalization process: in Old English, verb-particle constructions were present but had no preferential order. In Middle English post-verbal position for the particles gradually became the established norm. Through this period of decategorialization and loss of syntactic freedom, there was also an increase in construction types and in idiomatic meanings, all suggestive of a grammatical change (Thim, 2012:87–8). Indeed, the meaning overlap of particles in ambiguous contexts, where they could either indicate spatial movement or telic meaning (see Brinton & Closs Traugott, 2005:124), certainly contributed to this change. By the late 1800s, when the oldest individuals in the corpus were born, verb-particle combinations become more fixed and are known to have increased in frequency, particularly in speech-related text types; they are widely recognized as denoting aspectual meaning and show signs of lexicalization with the verb and the particle fusing to one lexical unit (i.e., phrasal verbs) (Brinton & Closs Traugott, 2005:124). While few studies have documented the alternation between the split and joined variants of these constructions in that time period, we can assume that, when Ontario dialects were established in the

mid-1800s, phrasal verbs were well entrenched in English grammar along with any social factors regulating their distribution and patterning. Therefore, we propose that the ongoing trend toward joined variants in Ontario English is diagnostic of its advanced stage development as part of the longitudinal grammaticalization of English prepositions into aspectualizing particles in the English verb phrase (cf., Brinton, 1996:163ff.). Indeed, according to Brinton (1985:160), particles can mark telicity as well as iterative and continuative aspect on the bare verb, turning activity verbs into accomplishment phrasal verbs, even to the point where the aspectualizing particles can be added to verbs where the perfective meaning is already present, for example, *finish up* instead of simply *finish*. If this development was only due to grammaticalization, we would expect an increase of phrasal verbs expressing idiomatic or abstract meaning over time, something that has been reported for their diachronic development (Claridge, 2000:96). A closer look at the present data reveals, however, that idiomatic phrasal verbs do not increase in frequency in the twentieth century in apparent time but stay roughly the same throughout the century, see Figure 11.

Such a development demonstrates that grammaticalization of the particle by itself cannot account for the increase of joined variants. Instead, concomitant lexicalization processes seem to be at play as well. As proposed by Rodríguez-Puente (2019:118), it is only particles and not the whole phrasal verb that have undergone grammaticalization (losing spatial meaning and acquiring grammatical, i.e., aspectual, meaning), while verb-particle combinations have undergone lexicalization (see also Thim, 2012:84), “gain[ing] lexical content and los[ing] grammatical properties” (Rodríguez-Puente, 2019:112). This lexicalization has brought with it not only a fusion of verb and particle but also a shift in meaning from compositional to more abstract, idiomatic meaning, similarly to the proposed grammaticalization process (Rodríguez-Puente, 2019:114). The fact that idiomatic uses of the verbs in our data do not increase in frequency, even though lexicalization (and grammaticalization) would predict it, might be explained by the comprehensive perspective we have taken here—that is, sampling multiple verbs—and with the gradual nature of the trend toward idiomaticity, which does not affect all verbs at the same time. Explaining the increase in joined variants as another phase in an ongoing lexicalization process finds support in related parts of the grammar (see Hundt & Zehentner [accepted], who have shown that prepositional phrases have become more integrated into the verb phrase during Early Modern English).

However, such systematic pressures inherent to the linguistic system can only partially explain our results. Social correlates such as speakers’ community and occupation also turned out to influence this alternation in Ontario. Why should only those two factors, but not gender or education, play a role? Sociolinguists have long claimed that syntactic variables are less likely than phonological or lexical variables to bear social meaning or show social stratification due to their abstract nature (see Cheshire, 2003:245; Levon & Buchstaller, 2015:320; also Lavandera, 1978). In contrast to phonological variables, which do not carry any propositional meaning, morphosyntactic or lexical variables can arguably

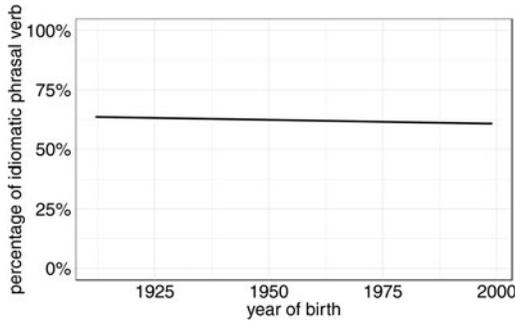


FIGURE 11. Proportion of idiomatic phrasal verbs by year of birth (regardless of variant).

express different functions and meanings depending on the context (see Tagliamonte, 2012:206). Studies that did consider morphosyntactic or lexical variables were thus often concerned with the alternation between a standard and a non-standard form that have clear social functions. With regard to language-external factors on the alternation between two standard forms, as in the present case, research is still lacking. The few studies that exist have mainly taken a broader perspective, focusing on style or regional background of the speaker (e.g., Bresnan & Hay, 2008; Röthlisberger, Grafmiller, & Szmrecsanyi, 2017) rather than factors relating to individuals' demographic background. With regard to other syntactic alternations in Ontario English, we know that, at the time of writing, the dative alternation shows no signs of education or occupation effects (Tagliamonte, 2014); while the preterit/present perfect alternation has shown a significant effect of education (Franco & Tagliamonte, to appear). Such mixed findings for so few variables in the same corpora do not yet point to a principled explanation; however, as more studies of syntactic alternations emerge from the Ontario Dialects Project, further comparisons will become possible. As to the broader question of why syntactic variables tend not to encode social meaning, one plausible explanation is that when multiple language-internal pressures are operating in tandem to reorder syntactic elements and evolve semantic meanings in grammar, they are less likely to be taken up to mark socially relevant aspects of communication. Certainly, as more contemporary studies of syntactic and semantic change are undertaken and greater capacity for cross-variety comparison is possible, further attention to this question in sociolinguistic theorizing will be possible. For now, the available evidence suggests that social and geographic factors can be involved in syntactic variables, in particular such influences as formal education or occupations, where attention to linguistic performance is required. In this regard, particle alternation in Ontario English seems to constitute an example of a socially stratified syntactic variation that refutes the so-called "Interface Principle" (see Levon & Buchstaller, 2015), meaning that speakers do indeed attribute social meaning to syntactic variables³ (see also the results in Kroch & Small [1978]). The degree of this attribution is likely to depend on the syntactic variable, its social meaning, and the nature of

communities where speakers live. Crucially, because the distribution of individuals with higher levels of education and professional occupations may not be parallel in all places nor in every generation equally, questions arising about the social influences on syntactic variables have added complexity.

It is also true that broad social categorizations, such as gender or occupation, might obscure underlying factors that impact these variables. The observed influence of occupation in our data could be due to stylistic shifts within the spoken conversations from formal discussions of, say, politics to less formal narratives of personal experience. Some studies on particle placement, for instance, have shown a higher frequency of the split variant in spoken discourse with the joined variant being more frequent in written discourse (Cappelle, 2006:8). What is more, Cappelle (2006:9), referring to Bolinger (1971:57, fn. 8), related the split order to colloquial, informal language. A similar correlation between the split variant and informal, colloquial language has been reported in Kroch and Small's (1978) study (see Kroch & Small, 1978:48–49). Both Grafmiller and Szmrecsanyi (2018) and Kroch and Small (1978) showed that, in comparison, the joined variant is more frequent in formal (written) discourse than in more colloquial (spoken) one. With regard to Ontario English, we cannot confirm a correlation between the joined order and formal language, because our focus has been on broad social and linguistic factors rather than detailed analysis of style-shifting within the recorded vernacular conversations. However, a correlation between the joined variant and more formal writing ties in with our findings regarding the impact of individual's occupation, with white-collar workers and students—who, characteristically, use more standard and formal language—preferring the joined variant more than blue-collar workers.

How then can we explain the overall change toward a more formal variant in spoken language, especially against the well-described trend of colloquialization in twentieth century English (Mair, 2006:183)? A similar change toward a more formal variant has been previously explained by pressures of prescriptivism (see Hinrichs, Szmrecsanyi, & Bohmann, 2015), and the same pressures might be applicable in spoken language. Such pressures would indeed promote the joined variant, as shown by a small-scale judgment study in Kroch and Small (1978). They report that the majority of their thirty-two participants considered the joined variant as more “correct” than the split order. To the extent that such attitudes transform into normative practice, they would explain the rise of joined variant over time.

Besides potential prescriptive pressure, there are two other explanations for why we observe an increase of the joined variant not reported for spoken language by earlier studies. First, it might be possible that we would find even more joined variants in the more formal registers of Ontario English, which would relativize our results. The higher proportion of joined variants versus split variants in our spoken interviews is, however, consistent with previous research that aggregates data from multiple registers, for example, Grafmiller and Szmrecsanyi (2018). Second, our study offers a synchronic apparent time perspective of approximately one hundred years across the twentieth century, something that no

other studies have done. If the studies sampling data from ICE or the BNC had taken an apparent-time approach and controlled for speakers' year of birth, they too might have been able to document an increase of joined variants. It thus becomes incumbent upon other researchers with access to socially stratified data, and especially cross-register data, to not only test for the effect of speakers' year of birth but other speaker-related factors as well.

CONCLUSION

Corpus-linguistic approaches to analyzing particle placement have focused on the structure, function, and linguistic correlates of this alternation. In contrast, we have taken a variationist perspective, carefully circumscribing the data to alternating contexts only and employing data that is sociolinguistically stratified so as to discern broad social patterns and apparent time trends. This enabled us to focus on the social embedding of this syntactic variable. In so doing, we have uncovered a system that is still developing and seems to be driven not only by language-internal processes but also by social pressures in the workplace, as indicated by statistically significant correlations with white-collar occupations and urban settings. Our methodology has several drawbacks. First, comparing our results to previous findings has turned out to be problematic due to differences in the nature of the data, the methods for circumscribing the data and the lack of information on social factors, such as unknown distributions of speakers' years of birth. Second, by sampling data from vernacular spoken language, we could not offer a multi-genre comparative perspective. Hence, the extent to which joined versus split variants would pattern across registers in Ontario English and even over time in present-day English more generally is not yet known. Third, our study does not consider other formal options such as the Latin equivalents of the phrasal verb (e.g., *find out–discover*) or the bare verb of the phrasal verb (e.g., *find out–find*), whose role in the overall development could be explored in future work. Finally, even though we have touched upon lexical considerations, a thorough investigation of lexical constraints on particle placement is certainly warranted (see Deshors, 2016). These limits notwithstanding, our study demonstrates that this syntactic option in contemporary Ontario English remains in flux, driven at least in part by conditions in the social world.

ACKNOWLEDGMENTS

The first author expresses their gratitude to the Fonds voor Wetenschappelijk Onderzoek (Belgium) for a travel grant. The second author gratefully acknowledges the Social Sciences and Humanities Research Council of Canada (SSHRC) for research grants 2001-present. We are indebted to the many fieldworkers and employees of the U of T Variationist Sociolinguistics Lab who collected and transcribed the data from 2002-present and to Rebecca Roeder for

access to the Thunder Bay data. We would also like to thank two anonymous reviewers and the editors for providing helpful feedback and the copyeditors Sonya Trawick and David Kaufman for their thorough work. All remaining errors are our own.

NOTES

1. All examples indicate pseudonym, age at time of interview, and community.
2. While Cappelle (2006:8) included pronominal objects that categorically favor the split variant, Gries (2003) included only 403 phrasal verbs. Therefore, the actual proportion of alternating variants is unknown.
3. We are thankful to one reviewer for pointing this out to us.

REFERENCES

- Baayen, R. Harald. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bates, Douglas, Maechler, Martin, & Bolker, Benjamin M. (2013). *lme4: Linear mixed-effects models using Eigen and S4* (R Package Version 1.1-7). <http://cran.r-project.org/package=lme4>.
- Belsley, David A., Kuh, Edwin, & Welsch, Roy E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley and Sons.
- Bolinger, Dwight. (1971). *The phrasal verb in English*. Cambridge: Harvard University Press.
- Bresnan, Joan, & Hay, Jennifer. (2008). Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2):245–59.
- Brinton, Laurel J. (1985). Verb particles in English: Aspect or Aktionsart? *Studia Linguistica* 39 (2):157–68.
- _____. (1988). *The development of English aspectual systems: Aspectualizers and post-verbal particles*. Cambridge: Cambridge University Press.
- _____. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter.
- Brinton, Laurel J., & Closs Traugott, Elizabeth. (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Brown, David West, & Palmer, Chris C. (2015). The phrasal verb in American English: Using corpora to track down historical trends in particle distribution, register variation, and noun collocations. In M. Adams, L. J. Brinton & R. D. Fulk (Eds.), *Studies in the history of the English language VI: Evidence and method in histories of English*. Berlin: Mouton de Gruyter. 71–97.
- Cappelle, Bert. (2009). Contextual cues for particle placement: Multiplicity, motivation, modeling. In A. Bergs & G. Diewald (Eds.), *Context and constructions*. Amsterdam: John Benjamins. 145–92.
- _____. (2006). Particle placement and the case for “allostructions.” In D. Schönefeld (Ed.), *Constructions all over: Case studies and theoretical implications*, vol. 1, <http://elanguage.net/journals/index.php/constructions>. 1–28.
- Cheshire, Jenny. (2003). Social dimensions of syntactic variation: The case of *when* clauses. In D. Britain & J. Cheshire (Eds.), *Social dialectology: In honour of Peter Trudgill*. Amsterdam: John Benjamins. 245–61.
- Claridge, Claudia. (2000). *Multi-word verbs in Early Modern English: A corpus-based study*. Amsterdam: Rodopi.
- Dehé, Nicole. (2002). *Particle verbs in English: Syntax, information structure, and intonation*. Amsterdam: John Benjamins.
- Denison, David. (1985). The origins of completive ‘up’ in English. *Neuphilologische Mitteilungen* 86(1):37–61.
- Deshors, Sandra C. (2016). Inside phrasal verb constructions: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Journal of Learner Corpus Research* 2(1):1–30.
- Elenbaas, Marion. (2013). Motivations for particle verb word order in Middle and Early Modern English. *English Language and Linguistics* 17(3):489–511.
- Franco, Karlien, & Tagliamonte, Sali A. (in press). Interesting fellow or tough old bird? 3rd person male referents in Ontario. *American Speech*.
- _____. (to appear). The preterit/present perfect alternation in Ontario English. February 2020.
- Geeraerts, Dirk, Kristiansen, Gitte, & Peirsman, Yves. (2010). Introduction: Advances in Cognitive Sociolinguistics. In D. Geeraerts, G. Kristiansen & Y. Peirsman (Eds.), *Advances in Cognitive*

- Sociolinguistics* (Cognitive Linguistics Research [CLR]). Berlin/New York: Mouton de Gruyter. 1–19.
- Gelman, Andrew. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27(15):2865–73.
- _____, & Hill, Jennifer. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Grafmiller, Jason, & Szmrecsanyi, Benedikt. (2018). Mapping out particle placement in varieties of English: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3):385–412.
- Gries, Stefan Th. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. New York: Continuum Press.
- Haddican, Bill, & Johnson, Daniel Ezra. (2012). Effects on the particle verb alternation across English dialects. *University of Pennsylvania Working Papers in Linguistics* 18(2):31–40.
- Hernández, Nuria. (2006). User's Guide to FRED (Freiburg English Dialect Corpus). MS. http://www.freidok.uni-freiburg.de/volltexte/2489/pdf/Userguide_neu.pdf (September 22, 2016).
- Hiltunen, Risto. (1983). *The decline of the prefixes and the beginnings of the English phrasal verb*. Turku: Turun Yliopisto.
- Hinrichs, Lars, Szmrecsanyi, Benedikt, & Bohmann, Axel. (2015). Which-hunting and the Standard English relative clause. *Language* 91(4):806–36.
- Hosmer, David, & Lemeshow, Stanley. (2000). *Applied logistic regression*. New York: Wiley.
- Hothorn, Torsten, & Zeileis, Achim. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16:3905–9.
- Hundt, Marianne, & Zehentner, Eva. (accepted). Prepositions in Early Modern English argument structure and beyond.
- Ishizaki, Yasuaki. (2012). A usage-based analysis of phrasal verbs in Early and Late Modern English. *English Language and Linguistics* 16(2):241–60.
- Jankowski, Bridget, & Tagliamonte, Sali A. (2020). Urban-rural dimensions to variable *-body/-one*: The case of Ontario, Canada. In A. Ziegler, S. Edler, N. Kleczkowski & G. Oberdorfer (Eds.), *Urban matters: Current approaches of international sociolinguistic research*. Amsterdam: John Benjamins.
- Kroch, Anthony, & Small, Cathy. (1978). Grammatical ideology and its effect on speech. In D. Sankoff (Ed.), *Linguistic variation: Models and methods*. New York: Academic Press. 45–55.
- Labov, William. (2007). Transmission and diffusion. *Language* 83(2):344–87.
- Lavandera, Beatriz. (1978). Where does the sociolinguistic variable stop? *Language in Society* 7(2):171–83.
- Levon, Erez, & Buchstaller, Isabelle. (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change* 27:319–48.
- Levshina, Natalia. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.
- Lohse, Barbara, Hawkins, John A., & Wasow, Thomas. (2004). Domain minimization in English verb-particle constructions. *Language* 80(2):238–61.
- Lutz, Angelika. (1997). Sound change, word formation and the lexicon: The history of the English prefix verbs. *English Studies* 78(3):258–90.
- Mair, Christian. (2006). *Twentieth-century English: History, variation, and standardization*. Cambridge: Cambridge University Press.
- Menard, Scott. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Thousand Oaks/London: SAGE Publications.
- Pinheiro, José C., & Bates, Douglas M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rodríguez-Puente, Paula. (2016). Particle placement in Late Modern English and twentieth-century English: Morpho-syntactic variables. *Folia Linguistica Historica* 37:145–75.
- _____. (2017). Tracking down phrasal verbs in the spoken language of the past: Late Modern English in focus. *English Language and Linguistics* 21(1):69–97.
- _____. (2019). *The English phrasal verb, 1650–present: History, stylistic drifts, and lexicalisation*. Cambridge: Cambridge University Press.

- Röthlisberger, Melanie. (2020). Social constraints on syntactic variation: The role of gender in Jamaican English ditransitive constructions. In T. Bernaisch (Ed.), *Gender in world Englishes*. Cambridge: Cambridge University Press. 147–75.
- Röthlisberger, Melanie, Grafmiller, Jason, & Szmrecsanyi, Benedikt. (2017). Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4):673–710.
- Schneider, Edgar. (2004). How to trace structural nativization: Particle verbs in world Englishes. *World Englishes* 23:227–49.
- Szmrecsanyi, Benedikt. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1):113–50.
- Szmrecsanyi, Benedikt, Grafmiller, Jason, Heller, Benedikt, & Röthlisberger, Melanie. (2016). Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2):109–37.
- Taeldeman, Johan. (2005). The influence of urban centres on the spatial diffusion of dialect phenomena. In P. Auer, F. Hinskens, & P. Kerswil (Eds.), *Dialect change: Convergence and divergence in European languages*. Cambridge: Cambridge University Press. 263–83.
- Tagliamonte, Sali A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden and Oxford: Wiley-Blackwell.
- . (2014). A comparative sociolinguistic analysis of the dative alternation. In R. Torres-Cacoullous, N. Dion, & A. Lapierre (Eds.), *Linguistic variation: Confronting fact and theory*. London/New York: Routledge. 297–309.
- Thim, Stefan. (2006). Phrasal verbs in Late Middle and Early Modern English: Combinations with *back, down, forth, out* and *up*. In C. Dalton-Puffer, D. Kastovsky, N. Ritt, & H. Schendl (Eds.), *Syntax, style and grammatical norms: English from 1500-2000*. Bern: Peter Lang. 213–28.
- . (2012). *Phrasal Verbs: The English verb-particle construction and its history*. Berlin/New York: Mouton de Gruyter.

APPENDIX A

List of variable verbs used to extract particle placement.

act, add, ask, back, bail, bang, barf, beam, bear, beat, beef, belch, bend, bite, black, blame, blare, blast, block, blot, blow, blurt, bowl, break, brighten, bring, brush, bubble, build, bundle, burn, buy, call, calm, carry, carve, cast, chalk, change, charge, chase, chat, check, cheer, choke, chop, chuck, churn, clean, clear, close, clutter, cock, collect, conjure, connect, cook, cool, cordoned, cough, count, cover, cross, cull, curl, cut, dash, dig, dish, divide, dog, dole, draft, drag, draw, dress, drive, drop, drown, eat, edge, edit, fake, feather, feel, fend, ferret, fight, figure, fill, find, finger, finish, fire, fish, fit, fix, flag, flatten, flick, fling, flog, flush, fold, follow, force, free, gather, gear, get, give, gobble, grab, graft, gulp, gun, hack, hammer, hand, hang, haul, head, heat, heave, help, hew, hire, hit, hold, hook, invite, iron, jack, jam, jot, keep, kick, kill, knock, lace, lay, lead, leave, let, level, lie, lift, light, line, link, list, liven, load, lock, look, loosen, lop, lose, lower, lure, make, map, mark, match, mess, mix, mop, move, muck, mull, note, offer, open, opt, order, pack, paint, pass, paste, pat, patch, pay, peel, phase, phone, pick, pile, pin, piss, play, plot, point, poke, pop, pound, pour, press, print, prop, pull, push, put, rack, raise, rake, read, rent, ring, rip, roll, rough, round, rub, rule, run, saw, scale, scar, scoop, scrap, scrape, scream, screw, scrunch, seek, sell, send, serve, set, settle, shake, shed, shift, shoot, shore, shorten, shove, shovel, show, shrug, shut, sign, sing, single, siphon, size, skim, slash, slice, slide, slow, smooth, snap, snuff, sort, sound, spat, speed, spell, spew, split, spoon, spread, spruce, squeeze, stagger, stamp, start, steer, step, stick, stir, store, straighten, stretch, strike, string, strip, stub, stuff, stutter, suck, sum, swallow, swap, sweep, swipe, swirl, switch, swoop, take, talk, tear, thin, think, throw, tidy, tie, tilt, tip, toss, touch, toughen, trace, track, trip, trot, try, tuck, tug, turn, twink, type, use, usher, vote, wait, wake, warm, wash, wave, wear, weed, weigh, whip, whisk, win, wind, wipe, work, wrap, wreck, wrench, wrest, write

APPENDIX B

Raw counts of joined and split variant by year of birth.

Year of Birth	Number of joined variants	Number of split variants
1912	3	3
1913	4	3
1916	2	4
1917	10	3
1918	5	6
1919	65	65
1920	16	47
1921	40	37
1922	33	44
1923	3	4
1924	16	16
1925	13	20
1926	33	61
1927	46	72

Continued

Continued

Year of Birth	Number of joined variants	Number of split variants
1928	9	18
1929	27	28
1930	28	63
1931	11	5
1932	22	21
1933	29	64
1935	41	56
1936	15	19
1937	47	60
1938	30	23
1939	25	21
1940	31	36
1941	54	51
1942	7	13
1943	13	24
1944	33	30
1945	29	25
1946	20	7
1947	46	49
1948	28	49
1949	29	25
1950	27	15
1951	67	44
1952	68	59
1953	35	48
1954	31	32
1955	85	122
1956	2	1
1957	23	27
1958	23	29
1959	30	45
1960	31	36
1961	20	45
1962	25	26
1963	62	33
1964	10	18
1965	38	45
1966	14	31
1967	91	99
1968	48	29
1969	13	8
1970	52	43
1971	10	11
1972	15	12
1973	20	23
1974	31	48
1976	24	10
1977	18	13
1978	60	23
1979	52	32
1980	13	14
1981	29	33
1982	56	42

Continued

Continued

Year of Birth	Number of joined variants	Number of split variants
1983	59	44
1984	84	82
1985	140	66
1986	147	97
1987	51	39
1988	96	69
1989	31	45
1990	81	39
1991	124	94
1992	61	39
1993	73	35
1994	79	44
1995	56	41
1996	31	6
1997	12	6
1999	2	2

APPENDIX C

Raw counts of joined and split variant by length of direct object in number of letters.

Number of letters in direct object	Number of joined variants	Number of split variants
2	1	3
3	26	30
4	121	349
5	200	225
6	206	313
7	293	393
8	342	395
9	260	284
10	240	274
11	230	173
12	178	100
13	169	89
14	122	63
15	107	35
16	95	39
17	75	37
18	75	26
19	60	17
20	46	18
21	40	15
22	33	8
23	33	8
24	21	5
25	21	6
26	15	3

Continued

Continued

Number of letters in direct object	Number of joined variants	Number of split variants
27	23	1
28	17	1
29	10	2
30	8	0
31	6	0
32	12	0
33	6	2
34	4	1
35	4	0
36	5	0
37	3	0
38	1	0
39	3	0
40	2	1

APPENDIX D

Raw counts of joined and split variant by age group and community.

People born	Number of joined variants	Number of split variants
Toronto		
before 1937	175	174
1938-1974	506	448
after 1975	802	450
North Bay		
before 1937	10	22
1938-1974	110	102
after 1975	101	94
Temiskaming Shores		
before 1937	117	230
1938-1974	175	254
after 1975	143	111
Kirkland Lake		
before 1937	174	241
1938-1974	186	250
after 1975	156	154
Timmins		
before 1937	37	41
1938-1974	132	120
after 1975	108	65
Thunder Bay		
before 1937	5	11
1938-1974	107	108
after 1975	69	41

APPENDIX E

Mixed-effects logistic regression of data without students (only fixed effects shown).

Factor	Level of categorical factor	<i>n</i>	Rel. Freq. joined %	β	SE	<i>p</i> -value
(INTERCEPT)		3887	47.6	-0.526	0.358	0.142
L LENGTH		3887		1.676	0.102	<0.001
IDIOMATICITY	compositional	1446	38.0		Reference level	
	noncompositional	2441	53.3	0.137	0.103	0.182
AGE GROUP	before 1937	1038	42.5			
	1938-1974 (compared to “before 1937”)	2143	47.6	0.101	0.182	0.579
	after 1975 (compared to “1938-1974”)	706	55.4	0.694	0.194	<0.001
PLACE	Toronto	1449	54.0	0.244	0.095	0.010
	North Bay	260	52.7	0.296	0.171	0.083
	Temiskaming Shores	888	40.4	-0.337	0.112	0.003
	Kirkland Lake	858	42.2	-0.235	0.117	0.046
	Timmins	207	52.7	0.222	0.186	0.281
	Thunder Bay	225	45.8	-0.169	0.181	0.350
EDUCATION	non	1575	45.6		Reference level	
	yes	2312	49.0	-0.037	0.127	0.770
GENDER	male	2106	46.6		Reference level	
	female	1781	48.8	0.008	0.107	0.938
OCCUPATION	blue-collar	1528	45.5		Reference level	
	white-collar	2359	49.0	0.303	0.129	0.018
INTERACTIONS						
AGE GROUP : IDIOMATICITY	before 1937 : compositional	413	32.4			
	before 1937 : noncompositional	625	49.1			
	1938-1974 (compared to “before 1937”) : compositional	780	37.1		Reference level for idiomaticity	
	1938-1974 (compared to “before 1937”) : noncompositional	1363	53.6	0.001	0.198	0.996
	after 1975 (compared to “1938-1974”) : compositional	253	49.8		Reference level for idiomaticity	
	after 1975 (compared to “1938-1974”) : noncompositional	453	58.5	-0.370	0.223	0.097