Marie Barking*, Ad Backus and Maria Mos

**Individual corpus data predict variation in judgments: testing the usage-based nature of mental representations in a language transfer setting**

https://doi.org/10.1515/cog-2021-0105
Received September 24, 2021; accepted June 18, 2022; published online August 10, 2022

**Abstract:** This study puts the usage-based assumption that our linguistic knowledge is based on usage to the test. To do so, we explore individual variation in speakers’ language use as established based on corpus data – both in terms of frequency of use (as a proxy for entrenchment) and productivity of use (as a proxy for schematization) – and link this variation to the same participants’ responses in an experimental judgment task. The empirical focus is on transfer by native German speakers living in the Netherlands, who oftentimes experience transfer from their second language Dutch to their native language German regarding the placement of prepositional phrases. The analyses show a large amount of variation in both the corpus and experimental data with a strong link across data types: individual speakers’ usage – but not the usage by other speakers – is a significant predictor for the speakers’ judgments. These results strongly suggest that, in line with a usage-based approach, variation between speakers in experimental tasks is linked to their variation in usage. At the same time, such usage-based predictions do not explain all of the variation, suggesting that other individual factors are also at play in such experimental tasks.

**Keywords:** corpus data; experimental data; individual variation; language transfer; usage-based linguistics

*Corresponding author: Marie Barking, Department of Culture Studies/Department of Communication and Cognition/Tilburg Center for Cognition and Communication, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands, E-mail: M.Barking@tilburguniversity.edu. https://orcid.org/0000-0001-8413-5153
Ad Backus, Department of Culture Studies, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands, E-mail: A.M.Backus@tilburguniversity.edu
Maria Mos, Department of Communication and Cognition/Tilburg Center for Cognition and Communication, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands, E-mail: Maria.Mos@tilburguniversity.edu

Open Access. © 2022 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
1 Introduction

One of the fundamental assumptions in usage-based approaches to language is that our linguistic knowledge is directly based on usage (Bybee 2006, 2011; Ibbotson 2013; Langacker 1987; Schmid 2016). However, directly testing this core assumption has proved to be difficult. As a result, complex methodological approaches have been developed that aim to link speakers’ current and previous language use in both language production and processing. For language production, this link has, for example, been tested with the so-called trace-back method, employing corpus data to test to what extent speakers’ language use at any given time is similar to what they have said or encountered before (see Lieven et al. [2003, 2009]; Quick et al. [2018a, 2018b] who applied this method in the domain of child language acquisition). Similarly, this link can be tested for language processing, oftentimes with the aim to gather more direct evidence about how usage affects linguistic representations. To do that, researchers have related measures of experimental tasks designed to tap into processing (e.g., reaction times, eye-tracking) to frequency patterns in corpus data. Such combinations of corpus and experimental data have generally shown that linguistic constructions are activated and processed more easily the more frequent they are in representative corpus data (Blumenthal-Dramé 2016; Schmid 2016). More recently, however, research on individual variation has raised the question to what extent amalgamated corpus data are really informative in that regard (Verhagen et al. 2018): people differ in their linguistic experiences, and, assuming that their mental language representations are directly based on these experiences (Bybee 2011; Langacker 1987, 2016; Schmid 2016; Tomasello 2000), so do these representations. As a result, we have to ask ourselves to what extent amalgamated corpus data are informative of those other speakers taking part in the experiments, and whether they thus allow for meaningful predictions regarding how given constructions are stored in the mental language representations of these other speakers.1

In this study, we collected corpus (see Section 2 Corpus study) and experimental data (see Section 3 Experimental study) from the same participants, making it possible to test how well the corpus data – both at the amalgamated level and at the individual level – predict the participants’ experimental responses. On a methodological level, this set-up provides new insights into the question how informative amalgamated corpus data are for individual speakers, as it directly contrasts the predictive power of the amalgamated and individual corpus data; on a theoretical level, it allows us to test the core usage-based assumption that our

---

1 We want to acknowledge here that in most cases, it is not feasible to collect corpus and experimental data from the same participants.
linguistic knowledge is based on usage. If that is the case, the individual variation that speakers show in their corpus data should be reflected in similar ways in their experimental responses. We particularly test this usage-based link for the cognitive mechanisms of entrenchment and schematization, which are two of the core mechanisms assumed to shape mental language representations – and as a result language processing and production – in usage-based approaches (Bybee 2011; Ibbotson 2013). Entrenchment refers to a process triggered by exposure: every time speakers hear or use a word or construction, it becomes more and more entrenched for them, resulting in all kinds of frequency effects such as faster processing times and more usage (see Section 1.1 for more details). Schematization is a process in which speakers notice similarities across the different linguistic units that they are exposed to and find the underlying patterns (e.g., English-speaking children over time noticing that main sentences in English generally have an SVO word order, Tomasello [2000]; see Section 1.2 for more details).

So far, much of the research on entrenchment and particularly on schematization has been done in the domain of language acquisition, with researchers trying to link the increasingly complex patterns that emerge in children’s language use to the increasingly schematic representations that these children have formed on the basis of their input. In this study, we test how these mechanisms can be applied to another linguistic domain by investigating entrenchment and schematization in contact-induced language change, where – just as in language acquisition – new complex patterns slowly emerge in the speakers’ language use. These patterns are oftentimes the result of language transfer (Barking et al. 2022; Jones 2005; Ribbert and Kuiken 2010; Schmid and Köpke 2017), and investigating them can therefore shine a light on how new patterns enter a language and how they become part of the language over time – both at the individual and group levels. Ultimately, this can inform our understanding about how the mechanisms of entrenchment and schematization operate, in general and in the case of bilingual speakers in particular, who might apply these mechanisms to structure their language in- and output across different languages.

In our investigation of these cognitive mechanisms, our focus is particularly on individual variation. We feel that such a perspective is needed for the following two reasons: first, more and more research has demonstrated how extensive individual variation is, both across and within speakers (see for example Barking et al. [2022]; Caldwell-Harris et al. [2012]; Verhagen et al. [2018]; Verhagen and Mos [2016] who showed individual variation for the cognitive mechanism of entrenchment; and see Dąbrowska [2004, 2020] for individual variation regarding the mechanism of schematization). Given this extent of variation, linguistic theories have to be able to account for it, which usage-based approaches do by making individual variation one of their core assumptions: speakers differ in their input –
which also differs over time for each speaker, in the cognitive abilities that they use to process that input, in regards to the attitudes they hold towards language, and likely in a multitude of other ways – and focusing only on group-level trends is therefore likely to mask a lot of this variation that is so informative about all of these factors that shape language use on an individual level. Secondly, focusing on individual variation can therefore present a great opportunity to test linguistic theories, especially those that take a usage-based perspective, which places so much emphasis on variation. As De Smet (2020: 4) argues, “explanations of linguistic phenomena that resort to psychological principles” like entrenchment and schematization should also – as these cognitive mechanisms operate in the mind of the individual speakers – “be able to make predictions about individual behavior”. This reasoning allows us to use individual behavior and particularly individual variation in that behavior as a testing ground for psycholinguistic theories.

In this study, we look at individual variation from these two perspectives: first, as something that is so extensive in the data that we simply cannot ignore it, and secondly, as a tool to test our usage-based assumptions. To do that, we collected corpus data from native German speakers living in the Netherlands, who frequently experience transfer from their second language Dutch to their native language German, and then asked the same participants to take part in an experimental judgment task, in which they were presented with German sentences that were similar to those that they had produced in the corpus data themselves. We explored individual variation in terms of entrenchment and schematization in both the corpus and judgment data and then tested how the variation was connected, that is, whether variation in speakers’ corpus data was linked to variation in their experimental responses.

In the following, we describe the mechanisms of entrenchment and schematization in more detail, particularly in regard to the usage-based nature of these mechanisms, argue how they might be related to the phenomenon of language transfer, and discuss open questions that the new set-up of this study – directly linking speakers’ usage and experimental data – might allow us to answer.

### 1.1 Entrenchment and language transfer

Entrenchment is an important notion in usage-based approaches. The idea is that every time that speakers use or hear a word or a construction, it becomes more and more entrenched in their mental language representation. As a result, these constructions are easier to activate (Blumenthal-Dramé 2016; Bybee 2011; Schmid 2016) and therefore also more likely to be selected in subsequent language use.
When it comes to individual variation, this link between usage and entrenchment is important: speakers lead different lives, are exposed to different input, and thus differ in the entrenchment of specific words and constructions. For instance, Verhagen et al. (2018) showed that job recruiters who are frequently exposed to word combinations such as the Dutch combination *goede contactuele eigenschappen* (‘good communication skills’) and *werving en selectie* (‘recruitment and selection’), which frequently appear in job ads, have a higher entrenchment level for these multi-word units compared to students who are not yet looking for a job, as evidenced by faster processing times, higher familiarity ratings, and more usage of these units. This strongly suggests that there is indeed a direct link between individual differences in linguistic experiences and in entrenchment levels (see also Caldwell-Harris et al. [2012] for a study on differences in entrenchment of multi-word units in prayers between orthodox and secular Jews).

Clearly, the process of entrenching words and constructions happens for each speaker individually, yet entrenchment has also been shown to help explain larger, community-wide language phenomena such as (contact-induced) language change (Backus 2014; Bybee 2011; De Smet 2016; Schmid 2020). To illustrate this, De Smet (2016) distinguishes between two processes that are important for language change to occur – innovation and replication – and links both of them to entrenchment. First, innovation might occur when a construction, for instance due to its high entrenchment level, starts to interfere with the selection of the standard way of saying something (normal replication; Croft [2000]), resulting in speakers using that other construction instead (altered replication; Croft [2000]). In a language contact situation, this might be due to interference from the other language (Backus 2015). For instance, in a previous study, we showed that native German speakers living in the Netherlands oftentimes experience transfer from their second language Dutch to German, and that the extent of transfer was linked to the frequency with which they used Dutch in their daily lives (Barking et al. 2022). Specifically, we looked at the placement of prepositional phrases within a sentence, which may occur before (1; middlefield) or after (2; postfield) the main verb in both Dutch and German (*gesehen* in (1) and (2) ‘seen’), but are placed more often in the postfield position in Dutch than in German (De Sutter and Van de Velde 2007; Van Oost et al. 2016). Transfer from Dutch to German is therefore likely to result in an increased use of the postfield position in German (see Fitch [2011] for a similar interpretation of the frequency increase of the postfield position in Pennsylvania German), and indeed we found that the more the native German speakers used Dutch – and as a result, the higher the entrenchment levels for the postfield position were likely to be for these speakers –, the more they also used this position in their German.
Secondly, replication plays an important role in the process of language change as well (De Smet 2016). Speakers for whom the normal replication is highly entrenched, are less likely to experience interference from a competing altered replication. For example, potential structures like English *stealer or German *Bauer (as a derivation of the verb bauen ‘build’) are not used due to the already entrenched words thief and Bauer ‘farmer’, respectively (Schmid 2007: 121), showing that entrenchment cannot only explain when an altered replication is used, but also when it is not.

The aim of the current study is to investigate these entrenchment effects at the level of words and constructions. As argued above, earlier work showed that there likely is a link between transfer and speakers’ language use, but it is important to note here that assessing speakers’ frequency of using their languages can only provide a very broad indication of their entrenchment levels. Speakers greatly differ in terms of the words and constructions that they encounter and use in their daily lives (Caldwell-Harris et al. 2012; Verhagen et al. 2018). Therefore, knowing only how often speakers use a language says very little about their entrenchment levels of the individual words and constructions within that language. This information is crucial as those entrenchment levels might ultimately determine which constructions are being transferred. This could be either on a construction-specific level, with some words and constructions being more frequent and thus more entrenched than others; or even on a construction-speaker-specific level, with individual differences between speakers in how frequently they use individual words and constructions.

To illustrate this, imagine a native German speaker living in the Netherlands who works as a job recruiter, similar to the participants in the study by Verhagen et al. (2018). This speaker is likely to be exposed to many Dutch job ads and thus to the word combinations that are typical of these texts like goede contactuele eigenschappen (‘good communication skills’). As a result, these Dutch combinations are
likely to be highly entrenched for her, maybe much more so than the German equivalents. Consequently, when she wants to express a similar meaning in German, the Dutch combinations might interfere, which could lead to language transfer such as the speaker saying *gute kontaktuelle Eigenschaften* instead of a more conventional expression in German such as *gute Kommunikationsfähigkeiten*. For this speaker, we would expect transfer of these typical word combinations to occur much more frequently (a) compared to other domains, in which she frequently uses German, and (b) compared to other speakers who have a different occupation. Some anecdotal evidence also comes from the corpus that is used in this study, which contains texts written by German-Dutch bilinguals working in the holiday home rental industry. In the corpus, there are many transfer examples that are related to this work experience, such as *Fliegfeld* ‘flyfield’ instead of *Flughafen* ‘flyport’ based on Dutch *vliegveld* ‘airport’, *Annullierung* instead of *Stornierung* ‘cancellation’ based on Dutch *annulering*, or *äusserlich fällig* ‘outwardly due’ based on Dutch *uiterlijk* meaning both ‘outwardly’ and ‘at the latest’. These examples strongly suggest that transfer is indeed influenced by entrenchment at a construction-speaker-specific level.

The first aim of this study is to investigate this link between transfer and entrenchment in more detail, both at the level of individual constructions and at the construction-speaker-specific level (see also Section 1.3), thereby providing new insights into the question how usage and entrenchment are related and how that in turn relates to the phenomenon of language transfer.

### 1.2 Schematization and language transfer

The second cognitive mechanism that this study focuses on is the process of schematization. According to the usage-based approach, speakers store linguistic input as exemplars in their mental language representation (Bybee 2006, 2011). These representations are not random, in the sense that speakers store exemplars that they perceive as similar closely together (Bybee 2011; Croft 2000; Goldberg 1995; Langacker 1987). They then generalize over these exemplars and thereby form increasingly schematic patterns (Bybee 2011; Ibbotson 2013; Schmid 2016; Tomasello 2000). As a result, linguistic knowledge is assumed to be “more complex than just a collection of words and a set of abstract, fully schematic, grammatical patterns” (Quick et al. 2021: 6): people store expressions as fixed multi-word units, as partially schematic constructions (i.e., “lexical-syntactic hybrids” with open and fixed slots; Quick et al. [2021: 4]), and as fully schematic patterns.

This conceptualization has been shown to be very helpful in understanding different patterns in language use (Backus et al. 2011; Bybee 2011; Doğruöz and
Backus 2009; Noël and Colleman 2018; Quick et al. 2018a, 2018b, 2021; Schmid 2016, 2018; Schönefeld 2015; Verhagen et al. 2018), and it might help to better understand speakers’ transfer patterns as well. To illustrate this point, we return to our example contact situation of native German speakers living in the Netherlands and briefly introduce another construction that has been shown to be frequently transferred by these speakers: the use of the Dutch complementizer *om* ‘to’ (Brons-Albert 1992, 1994; Ribbert and Kuiken 2010). In German, this complementizer is used only at the beginning of subordinate clauses that specify a purpose or goal. In Dutch, on the other hand, it can be used in other contexts as well, and research has shown that native German speakers living in the Netherlands frequently start to use the complementizer also in these contexts in their German. Additionally, Ribbert and Kuiken (2010) showed that in a grammaticality judgment task, these speakers – but not native German speakers not in contact with Dutch – also oftentimes accept the use of the complementizer in these new contexts, and they argue that this result suggests that “the patterns for the use of the German and the Dutch complementizer might have merged” for these speakers (2010: 46).

Such a high-level merging is oftentimes assumed in cases like this (see for example Seliger [1991: 173], who argues that this is a common process “in which a more complex and more narrowly distributed pattern of the L1 is replaced by a less complex and more widely distributed L2 pattern”). However, although this probably indeed happens in some cases, it is not the only possible explanation (and the same holds for many other cases of transfer and contact-induced change; Backus et al. 2011). For example, in our own study on the use of the complementizer *om* by native German speakers in the Netherlands, only some speakers showed evidence for such high-level merging, while many other speakers used the complementizer on a more idiosyncratic basis in these new contexts, likely based on partially schematic constructions or fixed multi-word units instead (see also Doğruöz [2014] who showed that changes in the pronoun use in Turkish by Dutch-Turkish bilinguals seem to be driven by specific multi-word units and not by grammatical patterns). Ultimately, these results challenge us to carefully examine whether speakers have formed a (new) schematic pattern or whether they have produced a given utterance, which might instantiate that pattern, based on partially schematic constructions or fixed multi-word units instead.

From a usage-based perspective, this raises the central question of how much of our linguistic knowledge is actually stored and used as highly schematic patterns – or in other words, how much syntax is there really in our mental language representation? –, or whether most of it is based on more lexical constructions instead (Dąbrowska 2020; Schmid 2018). Importantly, just as for the cognitive mechanism of entrenchment, this might differ both between individual speakers, with some speakers storing constructions at a more schematic level than others
(Dąbrowska 2004, 2012, 2020), between individual constructions, and maybe even at the construction-speaker-specific level. Given that from a usage-based perspective we assume schematic patterns to be the result of generalization over language input, we would expect this variation – to at least some degree – to also be usage-based, with speakers more likely to form generalizations the more their input triggers them to do so (e.g., based on type-token-ratios [TTR], see Bybee [2011]; De Smet [2020] for discussions on what triggers productivity). In this study, we aim to test this assumption by exploring individual variation in terms of schematization in both corpus data and experimental data, and then test to what extent the variation in the experiment can be accounted for by the variation in usage.

1.3 Research aim

The aim of this study is threefold: first to investigate the extent of variation between speakers and constructions in regards to the cognitive mechanisms of entrenchment and schematization, secondly to test whether this variation in speakers’ language use can account for variation in their experimental responses as a way to test to what extent mental representations can be said to be usage-based, third to contrast the results of such a speaker-specific analysis with an analysis of amalgamated corpus data to investigate whether amalgamated data are informative of the language use of individual speakers. To do that, we collected corpus data from native German speakers living in the Netherlands and analyzed these data in regard to the speakers’ placement of prepositional phrases in the postfield position. Subsequently, participants took part in a judgment task, in which they were presented with uses of the middlefield and postfield positions that are similar to or different from their own uses of these positions. In the following sections, we explain how the corpus data were analyzed in regards to the concepts of entrenchment and schematization. The reasoning for the design and analysis of the judgment task is further explained in Section 3.

1.3.1 Corpus analysis: entrenchment

As explained in the previous sections, entrenchment is a cognitive process triggered by frequency: the more frequent a certain linguistic construction (in this case: the use of the middlefield or postfield position), the more entrenched we assume this construction to be for a particular speaker. It is important to note that frequency in a corpus and entrenchment in the mind are likely not identical, however, “we cannot directly measure degrees of entrenchment” and therefore
“must rely on operational definitions that approximate the theoretical construct” (Stefanowitsch and Flach 2016: 101). To do that, this study measures (a) the frequency with which speakers use the postfield position in general and (b) the frequency with which they use specific lexical instances in that position. For the first measure, we counted the speakers’ number of middlefield and postfield uses and calculated a percentage of postfield use for each speaker. For the second measure, we performed collostructional analyses. These analyses provide measures of collostructional strength for a given item (Gries 2007; Stefanowitsch and Gries 2003), that is, measures of how strongly a given item is attracted to a particular slot in a construction (in this case: the middlefield or the postfield position). The higher the value, the higher the attraction, with values above 1.3 significant at $\alpha = 0.05$ and above 3.0 at $\alpha = 0.01$ (Stefanowitsch and Gries 2003). We performed these analyses for the verbs, prepositions, and verb-preposition-combinations in our corpus data to account for a range of different lexical instances within the postfield position. We want to stress here, however, that these analyses are by no means exhaustive: similar analyses could be done for units at different levels of specificity. For instance, rather than grouping all occurrences of the preposition in, it would be possible to differentiate between prepositional phrases with certain meanings (e.g., locative, temporal, etc.) to test whether those are attracted to a certain position (see further explanation in Section 2.2.2 below, in connection to the sentences in (6) which show that such representation levels are indeed likely to account for some of the variation in the corpus data).

1.3.2 Corpus analysis: schematization

Schematization refers to a process in which speakers notice similarities across different linguistic units and generalize over these similarities. As explained above, this can result in constructions at varying levels of schematicity in the speakers’ mental representations. In this study, we are interested in testing which level of schematicity speakers rely on when producing instances of the postfield position. For example, looking at a sentence as in (2) above, speakers might have produced that sentence by using a schematic representation of the postfield position, by having stored and activated the entire sentence as a fixed unit, or by using any representation in-between such a fully schematic and fully lexical one. To investigate this, we used the type-token-ratio of the postfield position, which is a measure of the lexical diversity within that position. The reasoning is that the more different lexical items speakers use in that position (i.e., the higher the ratio), the more likely it is that the speaker uses a more schematic representation for the postfield position. We (a) calculated the speakers’ overall type-token-ratio (i.e., the number of unique verbs used in the postfield position divided by the overall
number of postfield uses) and (b) their type-token-ratio for individual verbs (i.e., the number of unique prepositional phrases used in the postfield position in combination with a specific verb divided by the overall number of uses of that verb with the postfield position).

2 Corpus study

2.1 Method

2.1.1 Corpus

For this study, we compiled a corpus (total word count: 1,370,708 words) containing German e-mails written by eight native German speakers living in the Netherlands (average word count per speaker: 171,338.5 words, SD = 163,503, ranging from 13,760 to 513,654 words). These speakers worked in the customer service team of a holiday home rental company and wrote these e-mails between 2015 and 2017 as replies to customer questions. Therefore, the content was mostly about holiday home rental, such as information about houses, booking requests, payment information, etc. The data collection was approved by the company and by the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences. All speakers also signed a written consent form agreeing that their e-mails could be included in the corpus. On average, the speakers (M = 42.3 years, SD = 13.8; 8 females) had been 28.9 years old when they migrated to the Netherlands (SD = 7.00), ranging from 21 to 40 years. Most of them had moved to the Netherlands because their partner lived in the Netherlands or for work. All speakers indicated that they had learned most of their Dutch by living in the Netherlands. Two speakers took additional Dutch classes. At the time of the study, they had been living in the Netherlands for an average of 13.4 years (SD = 9.2), ranging from 2 to 26 years. Some of the speakers grew up bilingually, with German–Italian (speaker A), German–Swiss German (speaker E), and German–Tamil (speaker G).

2.1.2 Analysis

The e-mails were analyzed using a python script specifically written for the task of counting prepositional phrases in the middlefield and postfield positions in German texts, which performed with an overall accuracy of 94.8% on a set of 1,000 randomly retrieved prepositional phrases from the corpus (see Appendix A for a description of the program and its performance). For each prepositional
phrase (e.g., *sehen in der Stadt*, ‘see in the city’), we noted the speaker, the preposition (*in; N = 62*), the verb (*sehen; N = 789*), the prepositional phrase (*in der Stadt, N = 60,345*), and its position.

### 2.2 Results

In total, there were 44,903 uses of prepositional phrases in the middlefield position and 2,946 uses of prepositional phrases in the postfield position, which equals a percentage of postfield use of 6.16%. Figure 1 plots the percentages for the individual speakers, which range from 3.34 to 26.24%.

To measure the concept of schematization, we further zoomed in on these uses of the postfield position, specifically how productively each speaker used this position as indicated by their type-token-ratio. To calculate this type-token-ratio, we divided the number of unique verbs in the postfield position by the number of overall postfield uses. Importantly, the type-token-ratio has been shown to be heavily influenced by sample size, with larger samples generally leading to lower ratios (van Hout and Vermeer 2017). This is an issue for our dataset, as speakers greatly differed in their overall number of postfield uses, which ranged from only 58 up to 648 postfield uses. To solve this, we selected a random subset of 58 postfield uses (i.e., the lowest number of postfield uses by a single speaker) for each speaker, making it possible to directly compare the speakers’ ratios.

Figure 1 plots the resulting ratios against the speakers’ postfield percentage. It shows that speakers differed in regard to their lexical diversity in the postfield position, with ratios ranging from 0.45 to 0.83. A correlation analysis showed no relation between the ratio and the speakers’ postfield percentage (*r* = 0.42, *p* = .30), indicating that speakers can have similar percentages of postfield use by either using that position in combination with a wide variety of different verbs or with only a limited number of verbs instead (see Figure 1).

#### 2.2.1 Results entrenchment

To zoom in on the individual constructions that were used in the middlefield and postfield position, we conducted collostructional analyses of the verbs, prepositions, and the verb-preposition-combinations. To reiterate, these analyses provide measures of how strongly an item (in this case: a verb, preposition, or verb-preposition-combination) is attracted to a particular slot in a construction (in this case: the middlefield or the postfield position). The higher the value, the higher the attraction, with values above 1.3 significant at *α* = 0.05 and above 3.0 at *α* = 0.01 (Stefanowitsch and Gries 2003). We multiplied values by −1 for all constructions
attracted to the postfield position to differentiate between preferences for the middlefield and postfield position. The results are visualized in Figure 2, with each dot representing a different verb, preposition, or verb-preposition combination. Dots below the dashed line represent attraction to the postfield position and dots above the dashed line to the middlefield position. The figure shows that most values cluster around 0, indicating that these constructions have no preference for either position. At the same time, there are constructions that have a strong placement preference for one of the two positions.2

Figure 1: Percentage postfield use and type-token-ratio per speaker; speaker codes (A–H) are based on the percentage of postfield use.

———

2 There is some debate about how to best perform collostructional analyses (see for example Schmid and Küchenhoff [2013] for a detailed discussion). Their main criticism is that collostructional analyses are oftentimes wrongly interpreted: as the test is based on null hypothesis testing, values of collostructional strength provide a measure of the likelihood of the null hypothesis (i.e., that a certain lexical item is not attracted to a certain construction) and not about the strength of that attraction. This is especially important when comparing items with different frequencies, as sample size – as in any null hypothesis testing – greatly impacts collostructional strength. When interpreting the results in this study, it is therefore crucial to keep in mind that different items should not directly be compared in terms of their overall attraction strength but more so in terms of their position preference (i.e., whether the item falls below the dashed line indicating significant attraction to the postfield position or above the dashed line indicating significant attraction to the middlefield position).
Finally, we tested whether these placement preferences differ between speakers. To do that, we conducted collostructional analyses for each speaker separately. Figure 3 plots the results for the 50 most frequent verbs in the postfield position. The $x$-axis contains the 50 different verbs and the $y$-axis their collocutional strength for each individual speaker. Overall, the figure shows considerable variation between speakers, with many verbs significantly attracted to the middlefield position for some speakers and to the postfield position for others (i.e., verbs with dots both below and above the dashed lines, Figure 3).

2.2.2 Results schematization

Similarly to the entrenchment measures, we further zoomed in on the specific verbs used in the postfield position and calculated each verb’s type-token ratio. To do that, we counted the number of unique prepositional phrases that were used in the postfield position in combination with this verb. We calculated this type-token ratio (i.e., number of unique prepositional phrases in the postfield position used

---

**Figure 2**: Collostructional strength for the verbs, prepositions, and verb-preposition combinations in the amalgamated corpus data; negative values indicate a preference for the postfield and positive values for the middlefield position; dashed line at $-1.3$ and $1.3$ which indicate significance at $\alpha = 0.05$. 

---
with a specific verb divided by the overall number of postfield uses of the verb) for each speaker separately.

Figure 4 plots the type-token-ratios for each of the 50 most frequent verbs in the postfield position (i.e., the verbs also shown in Figure 3) per individual speaker. Each dot represents the type-token-ratio for a specific verb as used by a specific speaker. This could result in a total of 50 verbs × 8 speakers = 400 different dots, but the number is much smaller for the following two reasons: (1) not all verbs were used by all speakers and, if they were used, some verbs were not or only very infrequently used in combination with the postfield position; (2) in order to directly compare the different ratios, we created equally-sized subsets of postfield uses. These subsets consisted of 10 uses of a specific verb by a specific speaker in the postfield position, which allowed us to calculate the type-token-ratios of a sufficiently large number of verbs while still being able to directly compare the ratios. Any verbs that were used less than 10 times in the postfield position by a given speaker are thus not included in Figure 4, explaining the different numbers of dots in Figures 3 and 4.

**Figure 3:** Collostructional strength for the 50 most frequent verbs in the postfield position based on the amalgamated data; negative values indicate a preference for the postfield and positive values for the middlefield position; dashed line at −1.3 and 1.3 which indicate significance at $\alpha = 0.05$; see Figure 1 for the color-coding; see Appendix B for the verbs corresponding to these numbers.
The results show — just as was the case for the type-token-ratios for the individual speakers (see Figure 1) — that speakers use the postfield position in different ways, with some speakers using a certain verb in combination with a large set of different prepositional phrases and others with a limited number of fixed expressions (see Figure 4).

To further illustrate these different usage patterns, we discuss a number of verbs in more detail which showed interesting usage patterns in Figure 3 and/or Figure 4. We start with the verb *entnehmen* ‘to take from’ used by speaker G, which is the verb with the highest attraction to the postfield position (see frame A in Figure 3) and one of the lowest type-token ratios (see frame A in Figure 4). A closer look at speaker G’s corpus data revealed that she mainly used this verb in one very frequent fixed expression (see sentences [3a]–[3c]; prepositional phrases are marked in bold), showing that her high collostructional value for this verb is the result of this one fixed expression alone.


‘As we can see in our system, you have already returned from your vacation.’

‘As we can see in our system, the payment of €451.70 was due on 01.02.2017.’

c. *Wie wir entnehmen können aus unserem System, hat der Kunde zwei Buchungen für die gleiche Unterkunft.*

‘As we can see in our system, the customer has two bookings for the same accommodation.’

Other verbs show relatively high type-token-ratios, indicating that speakers used them with a variety of different prepositional phrases in the postfield position. For instance, sentences in (4) show examples of speakers using the verb *sein* ‘to be’ in the postfield position (see frame B in Figure 4). Note the different ways in which *sein* is used in these sentences (i.e., *sein* is used in combination with a past participle in [4a], in combination with an adjective in [4b], and in a deontic construction in [4c]), illustrating its versatility in the postfield construction.

\[(4) \quad a. \quad \text{Es bleibt dem Kunden unbenommen nachzuweisen, dass keine oder wesentlich geringere Kosten entstanden sind \textit{als mit den nachstehenden Pauschalen}.} \]

‘The customer is free to prove that no or substantially smaller costs have been made than the ones specified in the following rates.’

\[b. \quad \text{Er gibt hierin an, dass er einverstanden ist \textit{mit der Rechnung}.} \]

‘He says that he accepts the bill.’

\[c. \quad \ldots \text{\textit{dass der Reiseveranstalter unverzüglich zu informieren ist im Falle auftretender Mängel}} \]

‘… that the tour operator has to be informed immediately in the event of defects.’

Figure 4 also shows that individual speakers can differ in how productively they use certain verbs. For instance, example sentences (5a)–(5e; mistakes in German are corrected within brackets) all contain the verb *machen* ‘to make’ (see frame C in Figure 4). The first two sentences (5a) and (5b) are written by speaker A and illustrate how this speaker uses the verb in combination with many different prepositional phrases in the postfield position. The remaining sentences (5c)–(5e) are written by speaker C. Overall, this speaker uses the verb *machen* 99 times in her corpus data; 32 of which in combination with the postfield position, resulting in a highly significant attraction to the postfield position for her (collostructional value: –6.35). Closer inspection of her corpus data however shows that many of her postfield uses are uses of sentence (5c), showing that for this speaker – similarly to
the sentences in (3) – the high attraction to the postfield position is driven by one fixed expression. In other cases, she uses machen in combination with the middlefield position (see sentences [5d]–[5e]). Overall, this example thus shows that whereas speaker A uses [machen + postfield position] rather productively, speaker C does not.

(5)  


‘I looked for alternatives in the immediate vicinity of the original booking.’


‘Also if you make use of a cot.’

c. *Beachten Sie bitte, dass wenn Sie 20 Minuten nichts gemacht haben auf diese Seite [auf dieser Seite], Sie wieder ausgelogt [ausgeloggt] sind.*  

‘Please note that if you do not do anything for 20 min on this site you are logged out again.’

d. *Dann müssen wir uns mit dem Hauseigentümer in Verbindung setzen, ob wir im Fall des Kunden X eine Ausnahme bezüglich dieser Kosten machen können.*  

‘Then we have to contact the home owner, whether we can make an exception in the case of customer X regarding these costs.’

e. *Da Sie die Buchung jedoch bei Booking.com gemacht haben, […].*  

‘As you have made the booking through booking.com, […].’

Finally, we discuss example sentences that illustrate that there are also other representational levels at which speakers can store and use the postfield position. Examples in (6) show sentences which contain the verb informieren ‘to inform’ and are written by speaker F (see frame D in Figure 4). Based on these sentences, the following pattern emerges: speaker F places prepositional phrases with information about time (*am 01.01.2017* ‘on 01.01.2017’; *nach einigen Tagen* ‘after a few days’) in the middlefield position, prepositional phrases with information about the sender (*von Ihnen* ‘by you’) likewise in the middlefield position, but prepositional phrases with information about the object of complaint (*über die Problematik mit dem WIFI/TV* ‘about the problems with the Wi-Fi/TV’; *bei vor Ort auftretenden Problemen* ‘in the case of defects’) in the postfield position. This speaker worked in the department handling customer complaints and produced a large number of these sentences, suggesting that this overall pattern is likely to be highly entrenched for her. Overall, this example illustrates that it is not necessarily the verb (here: informieren) that triggers the postfield position, but also the content of prepositional phrases (here: information about the object of complaint), showing the different levels of specificity with which speakers may store the postfield position in their mental representation.
499

3 Interim conclusion

The corpus data revealed a number of interesting variation patterns: the eight native German speakers living in the Netherlands differ in the extent to which they use the postfield position in their German and in the number and kind of lexical items that they use in that position. We interpret this as an indication of considerable speaker variation in regard to the cognitive mechanisms of entrenchment and schematization. This finding is directly related to the first aim of this study, to explore the extent and kind of speaker variation in the corpus data. Having observed this variation between speakers in language use, we now turn to a more explicit, normative measure, i.e., grammaticality judgments. We test how this variation in language use is related to the speakers’ grammaticality judgments (aim 2) and whether this is the case for both the individual and the amalgamated corpus data (aim 3).

4 Experimental study

4.1 Method

4.1.1 Design

This experiment is a personalized grammaticality judgment task, in which the same participants who contributed to the corpus data judge German sentences
containing prepositional phrases in the middlefield or postfield position. These sentences are based on their own usage in the corpus data. We first present the design of this experiment, before discussing our analysis plan and hypotheses.

The experiment has a $2 \times 2$ design. The first independent variable is the position of the prepositional phrase in the stimuli sentence (binary: middlefield, postfield; manipulated in the stimuli sentences, see Figure 5); the second variable is the speakers’ preference for the combination (binary: middlefield, postfield; determined based on the corpus data). To determine this preference, we selected the ten verb-preposition-combinations that were most attracted to the postfield position and the 10 combinations most attracted to the middlefield position for each speaker (see Figure 5). This means that each speaker received a personalized set of stimuli based on her own usage in the corpus data. For each of the selected 20 verb-preposition-combinations, four sentences were formed, resulting in a total of 80 sentences per participant. In one sentence, the prepositional phrase occurred in the middlefield position, in the other sentence in the postfield position (see [7] for an example), and there were two filler sentences, which contained the same words but had a word order that was clearly incorrect in German. The prepositional phrases in the sentence were chosen in such a way that they resemble speakers’ actual uses as closely as possible.

Overall, there were thus two congruent conditions, in which position of the PP and speaker’s preference match (middlefield–middlefield, postfield–postfield) and two incongruent conditions, in which they do not match (middlefield–postfield, postfield–middlefield). The dependent variables were the grammaticality judgments (grammatical, ungrammatical) and the corresponding reaction times.
4.1.2 Hypotheses

In a first step, we compared the speakers’ judgments to those of a control group of native German speakers not in contact with Dutch (see Section 4.1.3). This allowed us to test whether the speakers’ judgments had changed due to contact with Dutch. We expected both the speakers with and without Dutch contact to accept instances of the middlefield and postfield position, as both positions are grammatical in standard German. However, the analysis revealed that while participants indeed generally accepted the middlefield position, they greatly differed in the extent to which they accepted the postfield position (see Section 4.2.1 and see Section 5 for a discussion).

In a second step, we tested whether we could explain some of the variation in judgments for the native speakers with Dutch contact based on their corpus data. Specifically, we expected participants (a) to be overall more likely to accept the postfield position the more frequently they use that position themselves (i.e., as indicated by their overall postfield percentage), (b) to be more likely to accept the postfield position in cases of verb-preposition-combinations that they themselves used in that position (i.e., as indicated by the results of their collostructional analysis), and (c) to be more likely to accept the postfield position the higher their type-token-ratio, that is, the higher their lexical diversity in that position. We expected the same pattern for the reaction times.

Finally, we tested whether speakers’ own language use is a better predictor for their responses than the language use of the other speakers. This set-up mimics experimental designs in which corpus data from speakers other than the ones participating in the experiment are used to predict experimental responses. It allows us to test to what extent amalgamated data are informative of the mental representations of individual speakers.
4.1.3 Participants

The same eight native German speakers living in the Netherlands who participated in the corpus study also took part in this experiment. Additionally, we asked them to contact participants for the control group themselves, for instance, by asking siblings or close friends matched for age, gender, and region in Germany. Some participants were not able to find a matched participant. In that case, they provided us with information about their hometown in Germany and we contacted participants in that city matched for age and gender ourselves. The control group consisted of 13 participants, as some participants provided more than one matched participant. On average, they were 47.8 years old (SD = 13.2). The native language of all the control participants was German, with four speakers indicating that they also use other languages regularly in their lives (i.e., Swiss German, English, Spanish, Tamil). For information about the native German speakers living in the Netherlands, see Section 2.1.1.

4.1.4 Procedure

Participants performed the experiment using the program PsyToolkit, which was specifically developed and has successfully been used to perform a wide range of experiments online (Stoet 2010, 2017). They first read instructions, informing them that they were going to be presented with written German sentences and that their task was to judge whether these sentences were correct German sentences by pressing one of two buttons. Participants first practiced this task with as many practice sentences as needed for them to judge five consecutive sentences correctly (with a maximum of 20 sentences). They then participated in the experiment which consisted of the 80 stimuli sentences constructed for that particular participant (see Figure 5). The sentences were presented in random order. Overall, the experiment took around 10 minutes to complete.

4.2 Results

4.2.1 Comparison speaker groups

In a first step, we compare the judgments by the native German speakers living in the Netherlands and the native German speakers not in contact with Dutch using a multi-level logistic regression model with the judgment as the outcome variable (binary outcome: sentence judged as acceptable, sentence judged as unacceptable) and the position of the prepositional phrase in the stimuli sentence (binary:
middlefield, postfield), and participant group (binary: native German speakers with and without Dutch contact) as predictor variables. As participants were exposed to different sets of stimuli sentences, we added only random slopes for participants (and not items).

The results of the logistic regression model comparing the judgments by the participants with and without Dutch contact are shown in Table 1 (marginal $R^2 = 0.16$, conditional $R^2 = 0.25^3$). Overall, the participant groups did not differ in their judgments. Participants were more likely to accept sentences with a PP in the middlefield position (acceptance = 89.2%) than in the postfield position (acceptance = 63.1%). This was the case for both the participants with and without Dutch contact.

Figure 6 plots this general pattern as well as the judgment pattern for each individual participant. In this figure, the native German speakers living in the Netherlands and their matched control participant(s) are displayed together. It shows that there is large individual variation in the participants’ judgements, even in the case of participants growing up and still living in the same area in Germany. For example, looking at the two matched control participants for speaker B, sisters who still lived in their hometown (see Figure 6), one participant rejected the postfield position in more than 60% of the cases, whereas the other only did so in around 15% of the cases.

4.2.2 Judgment data predicted by corpus data

To test whether the participants’ judgments depend on their usage patterns in the corpus data for the native German speakers living in the Netherlands, we ran a

---

Table 1: Multilevel logistic regression model predicting judgment (binary: 0 = grammatical, 1 = ungrammatical) based on PP position, participant group, and their interaction.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. error</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−2.32</td>
<td>0.28</td>
<td>−8.40</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Position</td>
<td>1.66</td>
<td>0.25</td>
<td>6.79</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Participant group</td>
<td>0.20</td>
<td>0.42</td>
<td>0.48</td>
<td>0.63</td>
</tr>
<tr>
<td>Position*participant group</td>
<td>0.005</td>
<td>0.40</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Random effects (std. dev.)</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 The marginal and conditional $R^2$ were calculated using the r.squaredGLMM function of the MuMln package in R. The marginal $R^2$ explains the proportion of variance explained by the fixed factors, the conditional $R^2$ describes the explained variation of both the fixed and random factors.
Figure 6: Overall and individual percentages of accepting and rejecting sentences with a PP in the middlefield and postfield position for the participants with and without Dutch contact; the graphs are ordered/color-coded according to Figure 1.
logistic regression model predicting the judgments based on the participants’ corpus measures. In the model, we predict the judgments (binary: acceptable, unacceptable) based on the position of the prepositional phrase in the stimuli sentence (binary: middlefield, postfield) and the preference of the combination based on both the data by other speakers and the individual corpus data (binary: middlefield, postfield). For instance, to predict the judgment by speaker G for a sentence with the combination *entnehmen aus* (‘to take from’), we used the placement preference for *entnehmen aus* (1) based on the usage of all speakers except speaker G and (2) based on the usage of speaker G only. We also added the speakers’ percentage of postfield use and their type-token-ratio to the model. Distractor items were excluded. We added random intercepts for the participants; random intercepts for items were not needed as each participant was exposed to a different set of stimuli.

Before discussing the results, we briefly describe how the preferences as determined by the individual data and the amalgamated data differed from each other. The preferences based on the individual corpus data were used to construct this experiment (i.e., 50% of the selected stimuli have a preference for the middlefield position and 50% for the postfield position based on the individual corpus data, see Figure 5). Compared to this baseline, the preference predictions based on the amalgamated corpus data differed in the following ways: for constructions with a preference for the middlefield position based on the individual corpus data, the amalgamated corpus showed a preference for the middlefield position in 85% of the cases and a preference for the postfield position in 11.23% of the cases; in 3.8% of the cases, no prediction based on the amalgamated data could be generated as the construction was never used by any of the other speakers. For constructions for which the individual speaker had a preference for the postfield position, there was a preference for the middlefield position in 29.2% of the cases based on the amalgamated data, a preference for the postfield position in 54.1% of the cases, and no use by the other speakers in 10.7% of the cases. Overall, these percentages show that the predictions between individual and amalgamated data can differ quite substantially.

Table 2 shows the results of the multilevel logistic regression model (marginal $R^2 = 0.22$, conditional $R^2 = 0.29$). Participants were more likely to reject uses of the postfield position (rejection = 35.8%) than uses of the middlefield position (rejection = 9.4%). There was no significant effect of preference or an interaction effect between position and preference when the preference was based on the amalgamated data, but there was a significant interaction effect when preferences were based on individual data. When presented with a sentence with a prepositional phrase in the postfield position, participants judged that sentence as ungrammatical in 43.8% of the cases when they themselves had a preference for the
field position for that sentence and in only 27.8% of the cases when they had a preference for the postfield position (see Figure 7). The participants’ percentage of postfield use or their TTR did not explain any additional variation in their judgments.

Figure 7 plots the judgment patterns for each individual speaker next to the general pattern. The pattern of less rejection of the postfield position for combinations with a postfield preference applies – to at least some degree – for every participant. At the same time, participants clearly differ in how they judge these sentences. Some speakers (speaker B, speaker C, speaker F) are quite accepting of postfield uses regardless of their own preferences. Other speakers very frequently reject those constructions (speaker A, speaker E, speaker H). Besides, while some speakers accept all or almost all postfield uses in the case of a preference for the postfield position, some speakers still reject quite a lot of those uses (speaker A, speaker E, speaker G), that is, they reject sentences in the judgment task that they themselves used in similar ways in their corpus data. Overall, it seems that speakers are stricter in their acceptance of postfield uses in the experiment than they are in their usage of these constructions in the corpus data, but that speakers differ in regard to how much stricter they are.

5 Discussion

The aim of this study was to investigate individual variation regarding the cognitive mechanisms of entrenchment and schematization and to test how these

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−3.75</td>
<td>1.50</td>
<td>−2.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Position PP</td>
<td>2.62</td>
<td>0.54</td>
<td>4.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Amalgamated data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference other speakers</td>
<td>0.15</td>
<td>0.65</td>
<td>0.23</td>
<td>0.82</td>
</tr>
<tr>
<td>Position*preference other speakers</td>
<td>0.06</td>
<td>0.78</td>
<td>0.07</td>
<td>0.94</td>
</tr>
<tr>
<td>Individual data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference individual</td>
<td>0.84</td>
<td>0.67</td>
<td>1.27</td>
<td>0.21</td>
</tr>
<tr>
<td>Position*preference individual</td>
<td>−1.88</td>
<td>0.79</td>
<td>−2.37</td>
<td>0.02</td>
</tr>
<tr>
<td>Percentage postfield use</td>
<td>−0.03</td>
<td>0.03</td>
<td>−0.81</td>
<td>0.42</td>
</tr>
<tr>
<td>Type-token-ratio postfield</td>
<td>1.81</td>
<td>2.27</td>
<td>0.80</td>
<td>0.42</td>
</tr>
<tr>
<td>Random effects (std. dev.)</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7: Overall and individual percentages of accepting and rejecting sentences with a PP in the middlefield and postfield position for verb-preposition-combinations with a preference for the middlefield or postfield position; the graphs are ordered/color-coded according to Figure 1.
mechanisms relate to the phenomenon of language transfer. We used a three-step approach: first, we systematically explored the variation regarding these two mechanisms in the corpus data to get a better understanding of the extent and kind of variation; secondly, we used the variation as a testing ground (De Smet 2020) for the core usage-based assumption that our linguistic knowledge is based on usage. To do that, the same speakers who contributed to the corpus data participated in a personalized judgment task, allowing us to investigate to what extent individual variation is linked across the two data types, that is, to what extent individual variation in the experiment is accounted for by the individual variation in the corpus data. Third, we completed the first two steps for both amalgamated and individual corpus data to test whether amalgamated data can be informative of the individual speakers’ language use.

Before discussing the results, we first want to make a few general remarks about the placement of prepositional phrases in Dutch and German. The different placement options have oftentimes been characterized as a form of free variation, with speakers being able to place prepositional phrases in the middlefield and postfield position interchangeably. However, many participants, including the ones not in contact with Dutch, judged a substantial minority of the postfield uses as ungrammatical, suggesting that the use of this position is much more restricted than has previously been assumed. Clearly, more research is needed on how this position can be used in German and why there seems to be so much variation regarding its acceptance. In the meantime, we need to remember for the following discussion that not every use of the postfield position by our native German speakers living in the Netherlands is necessarily the result of transfer (i.e., the postfield position is used in standard German as well, albeit infrequently) and that not every rejection of its use is necessarily a sign of them becoming stricter in their judgments (i.e., native German speakers not in contact with Dutch also reject some uses of the postfield position). Instead, what the results show is how speakers differ in that regard from each other and how these differences are related across their language use and experimental judgments.

For the first step – exploring the extent of individual variation in the corpus data –, we observed great variation at the speaker-level, at the construction-level, and at the construction-speaker level. Regarding the entrenchment measures (i.e., percentage of postfield use, collostructional analyses), speakers greatly differed in the frequency with which they used the postfield position, both in terms of the overall frequency of the postfield position (see Figure 1) and in the frequency of the lexical items they placed in that position (see Figure 3). Regarding the measures of schematization (i.e., type-token-ratio), results showed that speakers used the postfield position in different ways: some of them used the position rather productively, while others used it mostly in combination with only a limited
number of fixed expressions (Figure 1; see also Dąbrowska [2004, 2012 and 2020] for similar results regarding individual variation in schematization). Again, this also differed at a construction-speaker-specific level, with the same speaker using different constructions in different ways (see Figure 4). We want to stress here again that we do not claim to have investigated representations at every possible level of schematization. Instead, there are likely many more possible representations that speakers store and use, as illustrated by the sentences in (6); see also Schmid (2018) and Schönefeld (2015) for discussions on the analysis of schematicity based on corpus data.

In a second step, we investigated how this individual variation is connected across the corpus and experimental data for the native German speakers living in the Netherlands. To do that, we tested whether the variation in the corpus can account for the variation in the judgment task, showing to what extent the experimental variation is usage-based. Out of all the predictors that we used for the experimental data, only two turned out to be significant: first, the position of the prepositional phrase in the sentence, with speakers being overall more likely to reject uses of the postfield position, and second, the entrenchment level of the construction for that individual speaker, with speakers being less likely to reject the postfield position when they themselves had a preference for the postfield position for the lexical item used in the sentence.

This result has two important implications: first, it suggests that most speakers formed a representation of the postfield position at a highly schematic level, as they were generally more likely to reject uses of the postfield than of the middle-field position. Working from a usage-based perspective, this observation is not at all trivial: assuming that schematicity arises by speakers noticing similarities and then generalizing over these similarities in their input – a process which by definition starts with lexically specific utterances and only gradually becomes more and more schematic (Bybee 2006; Hilpert 2015; Schmid 2018; Schönefeld 2015; Tomasello 2000) – researchers have raised the question whether speakers actually store and use patterns at the most schematic level at all. For the postfield position, this seems to be the case. Secondly, the results illustrate that different levels of schematicity can be in competition with each other. In most cases, these different levels would result in the same output (e.g., see Quick et al. [2021:6]: “which of the three types of unit [lexically specific, partially schematic, fully schematic] are instantiated in a speaker’s actual utterance is, in general, impossible to say”), but not in this case: the general pattern of the postfield position is oftentimes rejected, whereas many of the frequent, and therefore likely to be entrenched, lexical instances are being accepted. In those cases, entrenchment of the specific instance thus seems to overrule the schematic pattern to reject the postfield position, ultimately resulting in the attested result that speakers are less likely to reject
sentences with a matching preference. Overall, this suggests that speakers store constructions at different levels of schematicity and that they attend to these different representations when making their judgments.

Despite this significant link between corpus and experimental data, the results of this study also showed that usage cannot explain all of the variation in the speakers’ judgments. In general, it seems that speakers are stricter in their judgments than they are in their usage in the corpus data, with many speakers rejecting postfield sentences even when there was a match in preference. This difference in strictness was not explained by the overall frequency or productivity with which speakers used the postfield position: some speakers who used the postfield position frequently and productively still often rejected it in the experimental task. One possible factor that might be able to explain this difference between speakers is the difference in setting. When writing the emails, participants did not know yet that they would be included in a study on their language use, however, they were likely to be very aware of that fact when participating in the experiment, perhaps making them more cautious about their language use. This mindset might have been triggered even more so by the specific set-up of the experiment (e.g., having to make a forced choice between grammatical and ungrammatical sentences, using a limited set of constructions potentially resulting in some participants noticing the contrast between prepositional phrases in the middlefield and postfield position in the stimuli sentences, etc.). Future research could explore to what extent the experimental setting indeed triggered such a different mindset and how that differed across participants, for instance, by asking them to reflect on their choices during the judgment task. For now, since different usage patterns resulted in different judgment patterns, our results suggest that the judgments were the result of both the automatic, unconscious processes of entrenchment and schematization as well as of other, more reflective processes, and that we need to account for both in our linguistic theories.

Regarding the third aim – comparing amalgamated and individual corpus data –, our results showed that it is only the speakers’ individual usage – and not the usage by other speakers – that predicts their responses in the judgment task. This was the case even for our participant group which can be characterized as relatively homogenous: although speakers came from different areas in Germany and had been living in the Netherlands for varying lengths of time, all of them were living in the same region of the Netherlands at the time of the study, worked for the same company, and the emails that they wrote for the corpus were all related to this shared work experience. The fact that so much variation was found, to the point that individualized predictions were so much better than predictions based on the language use of other speakers, clearly demonstrates the importance of individual variation. Importantly, we want to stress here that in most cases, collecting corpus
and experimental data from the same participants is simply not possible and that in those cases, using amalgamated corpus data from other speakers is the only available predictor of speakers’ language use. Additionally, the predictiveness of such amalgamated data might be a lot stronger than suggested by the results of this study, as our amalgamated corpus consisted of only seven speakers and is thus rather small. We therefore do not want to argue that researchers should abandon the use of amalgamated corpora as a predictor of language use altogether, but instead want to show with this study that, when using amalgamated corpus data in that way, we have to carefully think about what these data can and cannot tell us about the individual speakers’ language use.

5.1 Conclusion

In this study, we established a direct link between speakers’ own language use and their responses in a personalized experimental task. We used this approach in the domain of language contact, and it proved to be very informative in several regards. First, it showed that speakers can make use of transferred constructions in different ways – in terms of the frequency and productivity with which they use them and the lexical constructions that they use them with –, and that this greatly differs between speakers. Secondly, the approach showed that speakers’ language use and their grammaticality judgments are linked in intricate ways: on the one hand, their judgments largely depend on their usage, with speakers being more likely to accept constructions that they also used themselves; on the other hand, speakers also seem to make use of reflective processes during the experiment, resulting in them being much stricter in their judgments than in their usage. These are all important insights that help better understand the cognitive mechanisms of entrenchment and schematization in general and in their relation to language transfer in particular. At the same time, this methodological approach is of course not limited to the domain of language transfer. On the contrary, entrenchment and schematization – and thus also individual variation regarding these mechanisms (Dąbrowska 2020; Verhagen et al. 2018) – have been shown to play a key role in our language processing and production in general, and the approach of using individualized rather than amalgamated data to predict speakers’ responses in experimental tasks may therefore be a very fruitful way to go in future research, as a way to further explore individual variation in our language use.
Data availability statement

The data that support the findings of this study are available in DataverseNL: https://doi.org/10.34894/CQZAAT. This includes a file containing the extracted prepositional phrases from the corpus data and the data of the experimental judgment task. The program used to analyze the data is described in detail in Appendix A.

Appendix A: Automatic PP classification

Step 1: Finding all prepositional phrases in the text

In a first step, the program has to identify all prepositional phrases within the text. To do that, the program uses the German part-of-speech-tagging function and the dependency parsing function of the python library *spacy*. The first function assigns the part of speech to each word in a text (e.g., NOUN or VERB); the second function is a syntactic dependency parser which constructs a dependency tree of each sentence. Both functions – the tagger and parser – are trained on the TIGER corpus and German Wikipedia texts and they perform with a high accuracy of 97.15 and 89.75% respectively. To find all prepositional phrases in the text, the program first uses the POS-tags to find all prepositions (marked as ADP) in the text and looks up their corresponding noun phrases in the dependency tree.

Step 2: Splitting sentences

To split the text into sentences, the program uses the python *segtok* library, which is a pattern-based library specifically designed to split German texts into sentences. In a next step, the program uses the POS-tags to find all sentences which contain more than one finite verb (marked as VERB). If the sentence contains a subordinating conjunction (marked as SCONJ, such as ‘because’ or ‘when’) or a relative pronoun, the sentence is split at the comma and the part that contains the subordinating conjunction or the relative pronoun is classified as subordinate clause. Next, the program checks whether the sentence parts contain a conjunction (marked as CONJ, such as *und* ‘and’ and *oder* ‘or’). If this is the case, the sentence is

---

4 https://spacy.io/.
5 https://spacy.io/models/de.
split at the conjunction and both parts are classified as the type of clause that the sentence was also classified as before the split. Importantly, in order to be able to split the sentences correctly, the program relies on a comma to know where to split the sentence. However, commas are oftentimes (erroneously) missing in the text, especially in sentences that do not contain a subordinating conjunction. To minimize this problem, a fixed set of frequently occurring sentences and the correct way of splitting them into parts was specified manually (e.g., *wir hoffen* ‘we hope’ + subordinate clause; *wir bitten Sie* ‘we ask you’ + subordinate clause; *Sie haben die Möglichkeit* ‘you have the possibility’ + subordinate clause).

**Step 3: Identifying the position of the prepositional phrases**

In a last step, the program has to assign a label (no position, prefield, middlefield, or postfield position) to the prepositional phrases identified in Step 1. In main clauses, the program uses the POS-tags to check whether the sentence (or the previous sentence part for any main clause starting with a conjunction) contains an auxiliary (marked as AUX). If that is not the case, prepositional phrases are labelled as ‘no position’. If the sentence (or the previous sentence part) contains an auxiliary, the program compares the position of the auxiliary, the verb or adjective, and the prepositional phrase. The position is labelled as prefield if the prepositional phrase occurs before the auxiliary; as middlefield if the prepositional phrase occurs between the auxiliary and the verb (or adjective); and as postfield if it occurs after the verb (or adjective). In subordinate clauses, the program compares the position of the verb and the prepositional phrase. If the prepositional phrase occurs before the verb, the position is labelled as middlefield; if it occurs after the verb, as postfield.

**Step 4: Manual coding of verbs and prepositions**

After manual inspection, 200 ‘verbs’ (total instances = 686) and 19 ‘prepositions’ (total instances = 29) were excluded, as those had been incorrectly classified as verb or preposition by the *spacy* POS-tagger. All instances that included such a wrongly classified verb or preposition were excluded from further analysis. The remaining verbs and prepositions were manually corrected for spelling mistakes and different spellings (e.g., *außer, ausser*, ‘except for’). Different spellings of the preposition based on the gender of the following noun phrase (e.g., *zum* as the short form of *zu dem*; *zur* as the short form of *zu der*) were also grouped together.
Classification accuracy

To test the accuracy with which the program assigns the labels to the prepositional phrases, 1,000 random prepositional phrases were classified both manually and by the program. A comparison of these two different classifications showed that the program performed with an overall accuracy of 94.8%. Table 1 and A1 shows the accuracy scores within each category: no position, prefield, middlefield, and postfield position. The most common mistake of the program is to assign the label ‘no position’ to prepositional phrases that should be assigned a position. This is true for prepositional phrases in all three position (see Table A1).

Table A1: Labels in the test data and as classified by the program.

<table>
<thead>
<tr>
<th>Correct position</th>
<th>No position</th>
<th>Prefield</th>
<th>Middlefield</th>
<th>Postfield</th>
</tr>
</thead>
<tbody>
<tr>
<td>No position</td>
<td>391</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Prefield</td>
<td>3</td>
<td>84</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Middlefield</td>
<td>29</td>
<td>5</td>
<td>459</td>
<td>1</td>
</tr>
<tr>
<td>Postfield</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>

Based on these numbers, it is possible to calculate the program’s recall and precision for each of the four categories. The concept of recall describes what fraction of the prepositional phrases in one category (e.g., middlefield) is also classified as such. It is the number of times that the position is assigned to a prepositional phrase divided by the number of times the position actually occurs in the test data. The concept of precision describes how many of the prepositional phrases classified as one category are actually classified correctly. It is the number of times the position was correctly assigned to a prepositional phrase divided by the number of times that position was assigned overall. It is also possible to combine the scores of recall and precision into one value by calculating the so-called F1-measure. This measure is the harmonic mean of the two measures, which, for the case of two numbers, coincides with the square of the geometric mean divided by the arithmetic mean. Table A2 shows the recall, precision, and F1-measure for each of the four categories. Overall, these scores are very high and similar, which means that the program can give similarly accurate numbers of prepositional phrases for each of the four categories.
Table A2: Recall, precision, and $F_1$-measure for each position.

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No position assigned</td>
<td>98.73%</td>
<td>92.00%</td>
<td>95.25%</td>
</tr>
<tr>
<td>Prefield</td>
<td>94.28%</td>
<td>92.30%</td>
<td>93.28%</td>
</tr>
<tr>
<td>Middlefield</td>
<td>92.91%</td>
<td>98.92%</td>
<td>95.82%</td>
</tr>
<tr>
<td>Postfield</td>
<td>93.33%</td>
<td>93.33%</td>
<td>93.33%</td>
</tr>
</tbody>
</table>

Appendix B: Collostructional strength based on amalgamated corpus data

Table B3: Combinations with the highest collostructional strength for the postfield position.

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency</th>
<th>Collostructional strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>to be</td>
<td>3,127</td>
<td>-107.45</td>
</tr>
<tr>
<td>to take</td>
<td>300</td>
<td>-132.33</td>
</tr>
<tr>
<td>to mean</td>
<td>772</td>
<td>-16.28</td>
</tr>
<tr>
<td>to send</td>
<td>260</td>
<td>-51.02</td>
</tr>
<tr>
<td>to wish</td>
<td>39</td>
<td>-70.32</td>
</tr>
<tr>
<td>to start</td>
<td>14</td>
<td>-75.64</td>
</tr>
<tr>
<td>to receive</td>
<td>2,207</td>
<td>13.43</td>
</tr>
<tr>
<td>to make</td>
<td>215</td>
<td>-23.45</td>
</tr>
<tr>
<td>to log in</td>
<td>162</td>
<td>-26.28</td>
</tr>
<tr>
<td>to enter</td>
<td>24</td>
<td>-52.15</td>
</tr>
<tr>
<td>to book</td>
<td>1744</td>
<td>14.88</td>
</tr>
<tr>
<td>to pay</td>
<td>632</td>
<td>0.39</td>
</tr>
<tr>
<td>to arrange</td>
<td>66</td>
<td>-19.28</td>
</tr>
<tr>
<td>to take</td>
<td>239</td>
<td>-5.14</td>
</tr>
<tr>
<td>to record</td>
<td>1,149</td>
<td>8.29</td>
</tr>
<tr>
<td>to inform</td>
<td>587</td>
<td>0.99</td>
</tr>
<tr>
<td>to discuss</td>
<td>160</td>
<td>-6.43</td>
</tr>
<tr>
<td>to ask</td>
<td>49</td>
<td>-17.04</td>
</tr>
<tr>
<td>to found</td>
<td>0</td>
<td>-37.38</td>
</tr>
<tr>
<td>to pay</td>
<td>829</td>
<td>4.53</td>
</tr>
<tr>
<td>to wait</td>
<td>24</td>
<td>-21.52</td>
</tr>
<tr>
<td>to see</td>
<td>155</td>
<td>-4.67</td>
</tr>
<tr>
<td>to forward</td>
<td>502</td>
<td>1.05</td>
</tr>
<tr>
<td>to offer</td>
<td>241</td>
<td>-1.76</td>
</tr>
<tr>
<td>to ask</td>
<td>50</td>
<td>-11.63</td>
</tr>
<tr>
<td>to come</td>
<td>267</td>
<td>-0.84</td>
</tr>
<tr>
<td>to give</td>
<td>308</td>
<td>-0.47</td>
</tr>
</tbody>
</table>
Table B3: (continued)

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency</th>
<th>Frequency</th>
<th>Collostructional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>middlefield</td>
<td>postfield</td>
<td>strength</td>
</tr>
<tr>
<td>28 senden</td>
<td>667</td>
<td>24</td>
<td>4.03</td>
</tr>
<tr>
<td>29 zukommen</td>
<td>1,140</td>
<td>21</td>
<td>15.01</td>
</tr>
<tr>
<td>30 aendern</td>
<td>421</td>
<td>21</td>
<td>1.34</td>
</tr>
<tr>
<td>31 mitteilen</td>
<td>296</td>
<td>21</td>
<td>0.30</td>
</tr>
<tr>
<td>32 suchen</td>
<td>60</td>
<td>21</td>
<td>−7.12</td>
</tr>
<tr>
<td>33 bekommen</td>
<td>72</td>
<td>21</td>
<td>−6.03</td>
</tr>
<tr>
<td>34 anpassen</td>
<td>158</td>
<td>21</td>
<td>−1.97</td>
</tr>
<tr>
<td>35 gutschreiben</td>
<td>1,122</td>
<td>20</td>
<td>15.17</td>
</tr>
<tr>
<td>36 finden</td>
<td>298</td>
<td>20</td>
<td>0.39</td>
</tr>
<tr>
<td>37 entscheiden</td>
<td>47</td>
<td>20</td>
<td>−7.94</td>
</tr>
<tr>
<td>38 anlegen</td>
<td>188</td>
<td>20</td>
<td>−1.12</td>
</tr>
<tr>
<td>39 freuen</td>
<td>2</td>
<td>19</td>
<td>−19.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 stehen</td>
<td>251</td>
<td>18</td>
<td>0.28</td>
</tr>
<tr>
<td>41 entschuldigen</td>
<td>37</td>
<td>18</td>
<td>−7.92</td>
</tr>
<tr>
<td>42 erreichen</td>
<td>73</td>
<td>17</td>
<td>−3.95</td>
</tr>
<tr>
<td>43 angeben</td>
<td>201</td>
<td>17</td>
<td>−0.50</td>
</tr>
<tr>
<td>44 taetigen</td>
<td>304</td>
<td>16</td>
<td>0.94</td>
</tr>
<tr>
<td>45 erhohen</td>
<td>69</td>
<td>15</td>
<td>−3.29</td>
</tr>
<tr>
<td>46 hoeren</td>
<td>27</td>
<td>14</td>
<td>−6.56</td>
</tr>
<tr>
<td>47 liegen</td>
<td>54</td>
<td>14</td>
<td>−3.77</td>
</tr>
<tr>
<td>48 umbuchen</td>
<td>181</td>
<td>14</td>
<td>−0.34</td>
</tr>
<tr>
<td>49 reservieren</td>
<td>416</td>
<td>13</td>
<td>3.37</td>
</tr>
<tr>
<td>50 berechnen</td>
<td>113</td>
<td>13</td>
<td>−1.05</td>
</tr>
</tbody>
</table>

References


