The impact of co-occurrence and context on the prediction of long-distance separable prefixes

Daniel Walter and Brian MacWhinney

Carnegie Mellon University

Abstract

Current inquiry into language processing has placed an increased focus on people’s predictive capabilities in anticipating upcoming verbal elements. This study contributes to the literature by investigating the predictability of German verb prefixes, when they occur in sentence final position, often after a great deal of intervening material. Forty-nine L1 speakers of German were given a cloze-task to measure their ability to predict missing sentence final separable prefixes from a corpus of sentences taken from a German newspaper. The results show that German readers are able to accurately predict sentence-final prefixes and that accuracy is strongly correlated to cue strength between particular prefix-verb pairs, as well as the effect of other contextual clues. The discussion links this work to the implications for evolutionary advantages of prediction via alignment.

Keywords: prediction, psycholinguistics, corpus-linguistics, long-distance dependencies, co-occurrence

Recent studies have placed an increased focus on the ability of listeners to predict upcoming verbal elements. Functional approaches to language acquisition such as the Competition Model (Bates & MacWhinney, 1987; MacWhinney, 2014), connectionism (Baronchelli et al., 2013), and usage-based linguistics (Bybee, 2010; Ellis et al., 2015) all support the view of language skills as emerging from statistical patterns in the input (Erickson & Thiessen, 2015). Studies have also shown how frequency and statistical associations can facilitate word recognition (Balota & Spieler, 1999), syntactic processing (Spivey & Tanenhaus, 1998) and production fluency (Yoshimura & MacWhinney, 2007). Due to our shared experiences, the recurrence of distributional patterns is largely shared across members of a given social community. As a result, the underlying usage-based forces of cue strength, frequency, and validity play a major role in people’s ability to accurately predict upcoming words.

Within psycholinguistics, research related to speakers’ ability to predict upcoming words has typically focused on words in close proximity, often immediately next to one another. This study adds to the literature by investigating the predictability of verb prefixes in German, when they occur separated from the verb in sentence final position, often after a large amount of intervening material. These separable prefixes are members of a closed class of about 60 words (prepositions and locative adverbs) that are bound to particular verb stems. When the stem of the verb appears in the conjugated, second-position of a sentence, the prefix is placed in clause-final position. Often the meaning of the prefixed verb is not a simple semantic combination, but an idiosyncratic meaning only partially related to the combination. Thus, listeners cannot fully interpret the meaning of the verb until they reach the end of the clause, thereby needing to process a long-distance dependency between the two components of the prefix-verb complex. Since a given verb can have multiple separable prefixes, it is unlikely that a person could know with absolute certainty what the sentence final prefix should be. While the verb itself establishes the probability of particular prefixes, the additional sentence-internal context adds to and changes these probabilities, and therefore the probabilities for and against particular prefixes are continuously changing as more and more of the sentence is revealed. The results of the current study show that German native speakers are able to accurately predict sentence-final prefixes and that the accuracy of these predictions is strongly correlated to cue strength between particular prefix-verb pairs, as well as the effect of other contextual clues.

**Language Processing and Prediction**

An emphasis on the role of prediction in language processing has waxed and waned over the past fifty years. In the 1960s, researchers (Miller and Isard, 1963; Tulving and Gold, 1963) emphasized the extent to which listeners “generate hypotheses about upcoming words” (Van Patten and Luka, 2012). In this same vein, Goodman (1967) referred to reading as a “psycholinguistic guessing game.” However, in the 1980s, researchers (Forster, 1981) noted that, in general, the cloze probability for any given following word is so small that placing a general reliance on prediction would be a behavior with a “low payoff”. The impact of this critique held sway until the late 1990s, when advances in research technology, including eye-tracking and Event Related Potential (ERP) methods, brought the idea of prediction back into the picture (Van Petten & Luka, 2012).

A new wave of research into predictive linguistic behaviors has put a focus on the dynamic nature of prediction. Theories in language processing, such as the Competition Model (Bates & MacWhinney, 1987; Mitsugi & MacWhinney, 2014), neural network modeling (Elman, 1997; Christiansen & Chater, 1999) and surprisal theory (Hale, 2001) point to a continual online predictive process whereby probabilities are being calculated as new sentential elements are being introduced, and are being refined and corrected as more and more of a sentence is processed.

As a recent example of the way in which listeners use lexically specific data to predict upcoming words, Arai and Keller (2013) studied the effects of verb-specific information. In their study, the researchers used a visual world experiment (Tanenhaus & Spivey-Knowlton, 1996) modeled after Altmann and Kamide (1999), in which participants listen to sentences and view a scene that contains the target items and distractor items. Participants’ eye-movements were tracked to detect whether certain verbs, such as *eat*,were more likely to drive attention to food items in the picture than verbs like *move.* Like Altmann and Kamide (1999), Arai and Keller (2013) found that verb-specific information aided in participants’ abilities to narrow down and ultimately predict direct objects. In addition, Arai and Keller found that this verb-specific information can be used to avoid garden path ambiguity.

While visual world studies provide evidence that predictive behaviors do take place, one shortcoming of this type of study compared to real-world processing is the limited number of items from which one has to predict. Because visual world studies limit possible continuation items, it is difficult to tell whether predictive behaviors are limited to scenarios with ample context and limited choices, or if listeners make these predictions in a more general way as a basic part of language processing. While this question is still unanswered, these findings indicate that listeners and readers are not only integrating syntactic structure to predict upcoming words, but also word-specific semantic information.

ERP studies have provided further evidence for online lexical prediction. These studies have observed strong N400 responses both to semantically inappropriate words (Federmeier & Kutas, 2011; Kutas & Hillyard, 1980, 1983) and to words that fail to agree morphologically with other expected words (Tokowicz & MacWhinney, 2005; Dusias & Sagarra, 2007). These effects can be generated both by lexical and discourse contexts (Otten & Van Berkum, 2008).

The majority of the studies reviewed above have examined the prediction of open-class words (i.e. nouns, verbs, adjectives) from other open-class words. Problems regarding the complexity of predictions within this domain were at the focus of the critique of prediction models in the 1980s which is still alive today (Jackendoff, 2002). However, it is also possible that closed-class words that are closely associated with open-class words could play a major role in prediction. For example, in a sentence like *John picked \_\_\_\_ the book*, the word *up* immediately pops into mind, whereas *John put \_\_\_\_ the book*, elicits the opposition direction of *down*.

From the previous example, the connection between verb and prepositional choice is quite clear, but there is a problem. In these two sentences, the prepositions *up* and *down* are only obvious in print because we can see that the preposition occupies the slot between the verb and the direct object. Until we have processed the direct object, the sentences could also contain continuations such as *John put his hat on*, or *John picked out his favorite hat*. In German, however, the normal word order places direct objects before the postposed preposition. For example, *John picked up the book* in German becomes *John nahm das Buch auf* [John picked the book up]. Because of this, native speakers of German have more information from which they should be able to predict the lexical identity of sentence final separated prefixes.

**German Separable Prefix Verbs**

German sentences often rely on a structure known as the *Verbklammer*, or verbal bracket. This structure is composed of two related verbal elements that serve as a left and right bracket that surrounds other clausal elements. With clauses that contain multiple verbs, the conjugated verb makes up the left bracket and the other verbs make up the right bracket. Because of this, Germans must often wait a long time to hear the verb at the end of a clause (Thurmair, 1991). In sentences where there is only one verb, the conjugated verb is positioned in the left bracket next to its subject with the right bracket remaining empty. This causes an interesting effect with separable prefix verbs, which are composed of two elements: a verb stem and a prefix. If a separable prefix verb is the conjugated verb, then the prefix moves to the clause final position in the right bracket, leaving the conjugated verb stem in the left bracket position (Op. cit. pp. 194-195). The following example with the verb *vorstellen* illustrates how the single verb is used to fill in both the left and right brackets:

Ich stelle mich deinen Eltern vor

I place me your parents in front of

“I am introducing myself to your parents.”

The separable prefix verb *vorstellen* has a particular meaning that only becomes transparent once the sentence-final prefix *vor* appears well after the verb stem *stellen*. Note that *stellen* can also take other prefixes in combinations such as *einstellen*, *verstellen*, or *anstellen*. Thus, the prediction of the sentence-final preposition cannot be based solely on the identity of the stem.

Separable prefixes do not always separate from their verbs. In sentences with a modal verb, the verb and the prefix remain together in the right bracket at the end of the sentence.

Ich werde mich deinen Eltern vorstellen

I will me your parents introduce

“I will introduce myself to your parents.”

This is important because it shows that the probability of a verb-stem and separable prefix combination is driven by both its occurrence as a combined and separated verb. This will be a key aspect to keep in mind during the analysis.

Some combinations of verbs and separable prefixes have very specific meanings, as in the example of *vorstellen*, which could be analyzed compositionally as “in\_front\_of+place”, although its actual meaning is “introduce”. Others simply add compositionally to the core semantics of the main verb. For example, the verb *gehen*, [to go], can combine with other separable prefixes that do not really change the core meaning of the verb, but rather indicate in which direction one is going (e.g. *ausgehen* [to go out]*, eingehen* [to go in]*, losgehen* [to get going]*, hingehen* [to go to]).

Because of these various features, German separable prefix verbs provide an interesting opportunity to test whether the distribution of co-occurrences between specific verb-prefix pairings are strong enough to allow listeners to not only predict what syntactic argument is upcoming, but which specific lexical item is most likely to occur with a particular verb.

**Research Questions**

The research questions for the study are as follows:

1. Are German native speakers able to predict the sentence final separable prefixes used in examples taken from a corpus of authentic material?
2. What differences, if any, exist between the rates of successful prediction between items?
3. How does more or less context affect listeners’ abilities to predict sentence final separable prefixes?
4. If there are differences between items, what information can the corpus provide about why these differences appear?

**Methods**

**Participants**

Forty-nine first-language German speakers from the University of Bremen were recruited to participate in this study. All participants were at least eighteen years of age and were currently enrolled at the university.

**Design and Measures**

To test participants’ abilities to predict sentence-final separable prefixes a cloze task comprised of naturalistic data taken from a written corpus was chosen. The items are possible sentences that speakers of German could easily encounter in everyday life.

The cloze task consisted of 40 sentences containing separable prefix verbs with the verb in the second position and the separable prefix in the sentence final position. First, a list of all possible separable prefixes was compiled and searched for in the corpus. Sentences were selected at random from a corpus of the German newspaper *Rhein-Zeitung* from 1996 to 2011. A newspaper corpus was selected because of its authenticity and the fact that this is the type of written information to which adult speakers of German would have access. This corpus was accessed through the Cosmas II corpus analysis application provided by the Institute for German Language at the University of Mannheim.

In order to search the Cosmas II database for separable prefix verbs, the following command were used for each item:

&VERBSTEM /+s0 SEPARABLEPREFIX

e.g. &gehen /+s0 ein

This command searches for the verb stem in any conjugation and then the separable prefix occurring within the same sentence but not as a connected whole. This search resulted in sentences in which the conjugated verb appeared in the second position in the sentence and the separable prefixes occurred at the end. Some sentences that were found contained both words but they were not separable prefixes. For example, *ein* is also the indefinite article and the number “one” in German, so any instances where *ein* appeared with *gehen* but not used as a separable prefix were discarded. Four hundred sentences containing sentence final separable prefixes were identified and from this list forty items were selected at random.

Each item was divided into a more-context and less-context version. In the more-context version, the entire paragraph from the corpus up to and including the target sentence was used as the item. In the less-context version of the item, only the sentence in which the separable prefix was contained was used as the item.

A web-based online cloze task was created to administer the test to participants remotely. Each sentence was assigned an item number and retyped exactly as it had appeared in the newspaper, with the exception that the sentence final separable prefix was replaced by a text box. The items were chosen with equal probability from the two possible versions with either more or less context. Each participant was presented with an equal number of more and less context items.

**Procedures**

All participants were tested simultaneously in a lab setting and were monitored by an on-site administrator from the German university in which all participants were matriculated. Participants were given each of the 40 items one at a time, in random order, and randomly given either the more or less context item. As participants completed each item, the home-server collected and stored the data for each participant and item.

**Results**

Regarding research question 1, the results indicate a fairly normal distribution with a slight left-tail skew due to one participant’s performance. The distribution is shown in table 1. Overall, the mean accuracy among participants was *M =* 27.4 correct out of 40, with a standard deviation of *SD* = 4.1

INSERT FIGURE 1 HERE

From these descriptive statistics it is evident that this group of speakers was fairly successful at predicting the missing separable prefixes. Because of the online nature of the task, it was not possible to collect additional demographic information on each participant, which could have provided insight into why one participant faired so poorly in comparison to the other participants. One possible explanation is that the participant was not a native speaker of German or did not understand the instructions. Because this result is an extreme outlier, this participant’s data have been excluded from the remainder of the analysis. After removal of this participant, the group average increased to *M* = 28.02 out of 40 items (70.05%) with a standard deviation of 2.78.

Regarding research question 2, we can see that there are stark differences between items. While participants showed an overall ability to predict particular separable prefixes, the accuracy level of these predictions varied markedly across items. Figure 2 shows the percent correct in descending order left-to-right of each separable prefix verb.

INSERT FIGURE 2 HERE

Research question three asked whether differing amounts of context would affect the participants’ abilities to predict the sentence final particle. A t-test for between group differences for the items by more context (M= 17.03, SD = 8.55) versus less context (M = 17.08, SD = 8.78) variations revealed no statistical difference between groups (t = 0.026, df = 78, SE = 1.938, p = .980) which is unexpected based on previous research that has shown the importance of context in comprehension and prediction.

In order to address the fourth research question about what is happening for each verb, it is important to go back to the corpus to look for the factors that differentiate one verb from another. As a first step, all of the possible separable prefixes for each item stem were determined through the use of an online grammar tool (duden.de). For example, for the verb stem *stimmen* in item 39, which means “to be right” or “to be correct”, 17 possible prefixes were discovered. Each prefix changes the meaning of the word stem, as in *abstimmen*, which means “to vote”, or *zustimmen*, which means “to agree”. The number of co-occuring separable prefixes varied from one verb stem to the next, which can be seen in column four of table 1.

Once a list of all possible co-occurring separable prefixes was created for each verb stem, the frequency of each possible verb stem and separable prefix pair was investigated within the same written corpus used to select the items. In doing so, the proportion of the actual item could be seen in relation to the total uses of the verb stem by itself and with other separable prefixes. Figure 3 below shows the scatterplot of percent-correct by number of other co-occuring separable prefixes with the verb stem. There is significant negative correlation between the number of co-occuring verb-prefix pairs and the percent correctly predicted (R = -.382, p < .05), as is indicated later in Table 3.

INSERT FIGURE 3 HERE

Next, the corpus was also analyzed to see if there were any other frequently occurring words within the sentences. This search tracked both non-separated and separated uses for each verb.

&WHOLEVERB oder (&VERBSTEM /+s0 SEPARABLEPREFIX)

e.g. &eingehen oder (&gehen /+s0 ein)

This request pulled all sentences for each verb-prefix pair that were both joined, as in *eingehen*  as well as separated, as in *gehen* with *ein* at the end of the sentence.

In order to identify the number of other possible separable prefix pairings with the verb, Duden online (duden.de) and a Google search for separable prefixes for each verb stem was conducted. For example, the stem for item #11 *stellen* was found to have 14 possible prefix combinations, including *vorstellen*, *aufstellen*, *darstellen*, *ausstellen*, and so on. The total number of possible prefix combinations with each verb stem is listed in column four in Table 2 below. In addition to the number of total co-occuring prefixes with each verb stem, column 3 of Table 2 indicates what percentage of the total uses in the corpus were the combination of that particular verb stem and prefix pair. For example, in column three for item #4 *anbieten*, the number .9632indicates that, when the verb *bieten* was used with a separable prefix, it occurred with the prefix *an* 96.32% of the time.

As a further level of analysis, other closed-class words within the sentence were analyzed for relative frequency by running a concordance for each of them with the verb-prefix pairings. In four instances, an additional word was found to be predominantly associated (appearing in over 25 % of the token sentences) with the target item. For example, item #10 *eingehen* was found to occur with the accusative preposition *in* more frequently than with any of the other possible *gehen* + prefix possibilities. This is indicated in column five of Table 2 below.

The final two columns of Table 2 indicate the number of correct responses out of 49 and the percent correct by item. Results are organized from the most successfully predicted prefixes to the least successfully predicted.

Table 2: Statistics by item

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Item # | Item | % of all verb-prefix tokens | # of co-occurring prefixes | | Strong 2nd collocate | Number correct | %  correct |
| 4 | anbieten | 0.9632 | | 7 |  | 48 | 1 |
| 14 | vorlesen | 0.1583 | | 14 |  | 48 | 1 |
| 23 | umgehen | 0.1609 | | 15 |  | 48 | 1 |
| 10 | eingehen | 0.0857 | | 15 | accusative “in” | 48 | 1 |
| 13 | aussprechen | 0.2873 | | 16 |  | 48 | 1 |
| 17 | abriegeln | 0.9719 | | 3 |  | 48 | 1 |
| 33 | aussprechen | 0.2873 | | 16 |  | 47 | 0.979167 |
| 1 | eintreten | 0.0847 | | 13 | accusative “in” | 47 | 0.979167 |
| 26 | zurückgreifen | 0.0897 | | 14 |  | 47 | 0.979167 |
| 32 | eingreifen | 0.1944 | | 14 |  | 47 | 0.979167 |
| 34 | einstimmen | 0.2146 | | 6 |  | 46 | 0.958333 |
| 38 | abstimmen | 0.2051 | | 6 | über | 46 | 0.958333 |
| 6 | durchsetzen | 0.1414 | | 14 | sich | 46 | 0.958333 |
| 9 | ankündigen | 0.8880 | | 5 |  | 46 | 0.958333 |
| 28 | abspielen | 0.0634 | | 13 |  | 46 | 0.958333 |
| 25 | ausstellen | 0.0745 | | 14 |  | 45 | 0.9375 |
| 37 | aufräumen | 0.1850 | | 5 |  | 45 | 0.9375 |
| 20 | zurückkehren | 0.7123 | | 4 |  | 44 | 0.916667 |
| 22 | übergreifen | 0.0755 | | 8 |  | 43 | 0.895833 |
| 8 | vorkommen | 0.0458 | | 17 |  | 41 | 0.854167 |
| 11 | vorstellen | 0.2522 | | 14 |  | 41 | 0.854167 |
| 27 | wegfangen | 0.0015 | | 5 |  | 39 | 0.8125 |
| 31 | zusammenfallen | 0.0043 | | 22 |  | 39 | 0.8125 |
| 19 | zusammenschlieβen | 0.0289 | | 7 |  | 37 | 0.770833 |
| 35 | umsetzen | 0.1285 | | 14 |  | 36 | 0.75 |
| 5 | loslegen | 0.0251 | | 16 |  | 34 | 0.708333 |
| 7 | abgeben | 0.1089 | | 10 |  | 30 | 0.625 |
| 12 | vorstellen | 0.2522 | | 14 |  | 30 | 0.625 |
| 24 | lossagen | 0.0009 | | 21 |  | 27 | 0.5625 |
| 15 | mitnehmen | 0.0497 | | 20 |  | 26 | 0.541667 |
| 16 | zurückweisen | 0.0410 | | 15 |  | 18 | 0.375 |
| 36 | zusammenfinden | 0.0941 | | 11 |  | 18 | 0.375 |
| 18 | zusammensein | 0.0167 | | 25 |  | 17 | 0.354167 |
| 2 | zusammenarbeiten | 0.1063 | | 17 |  | 16 | 0.333333 |
| 39 | losprügeln | 0.0209 | | 11 |  | 11 | 0.229167 |
| 29 | einkommen | 0.0801 | | 17 |  | 2 | 0.041667 |
| 30 | anfügen | 0.1338 | | 7 |  | 2 | 0.041667 |
| 40 | weggehen | 0.0084 | | 15 |  | 2 | 0.041667 |
| 3 | überkosten | 0.0626 | | 19 |  | 0 | 0 |
| 21 | unterlegen | 0.0518 | | 14 |  | 0 | 0 |

Using a Pearson test for correlation for the variables of percent correct, the percent of the total verb-prefix pairings for that particular verb-prefix pairing versus the total tokens of all possible verb-prefix parings in the corpus, number of other possible verb-prefix parings, and presence of a strong second collocate, there is a clear relationship between the predictor variables and the outcome variable percent correct, as presented in Table 3

Table 3: Pearson Correlation of Percent Correct by Independent Variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Percent correct | | Percent of all verb-prefix tokens | Number of other prefixes | Strong second co-occurrer |
| Percent correct | 1.000 |  | |  |  |
| Percent of all verb-prefix tokens | .382\*\* | 1.000 | |  |  |
| Number of other prefixes | -.316\* | -.548\*\*\* | | 1.000 |  |
| Strong second co-occurrer | .270\* | -.073 | | -.053 | 1.000 |

\*p < .05, \*\*p < .01, \*\*\*p < .005

These test results indicate a significant, positive relationship between the percent correct and both the percentage of the total verb-separable prefix pairs in the corpus (*r* = .383, p = .007) and the presence of a second strong co-occurring word (*r* = .269, p=.046). There is also a significant, negative relationship between the percent correct of a particular item and the total number of possible verb-prefix pairings (*r* = -.318, p = .023).

**Discussion**

To understand the ways in which native speakers achieve long-distance prediction, we need to look at both overall accuracy and differences in accuracy levels across items. For research question 1, it is clear that native speakers of German are, in general, able to predict sentence-final separable prefixes. This ability shows astounding consistency between native speakers, with a standard deviation of only 2.78 items. For the overwhelming majority of items (30/40), participants were able to correctly predict the missing separable prefix over 50% of the time, despite the fact that these verb-stems have multiple other possible prefix pairings. This result provides an interesting insight into the predictability of different words based on word-class. Given these results, we can conclude that arguments against prediction based on the notion of a “low-payoff” do not apply to these types of words.

Research question 2 asks whether predictability is influence by the frequencies of individual verb-prefix pairings. Table 2 shows that there is a very different probability of success between items. Some items were correctly predicted 100% of the time, while others were never correctly predicted. This variability within this closed-class of words is significant because it indicates that there may be a complex set of forces driving predictability.

Research question 3 asks whether prediction can be improved through generation of a supportive discourse context (Bransford, Barclay, and Franks, 1972). Somewhat surprisingly, no such improvement was observed in these materials. This negative finding suggests that prediction of sentence final particles is driven more by lexical and syntactic factors than by top-down inferencing from the overall discourse context.

Research question 4 seeks to identify some of the underlying mechanisms that drive differential predictivity for particular verb-prefix combinations. From the current analysis, it appears that these probabilities are altered by the proportion of the frequency of the verb-prefix pair in relation to all other verb-prefix possibilities, the total number of alternate verb-prefix probabilities, and the presence of other sentence-internal cues, among other possible factors such as semantic relevance, or others not specifically investigated in this study.

First, the more frequently a separable prefix occurs with a verb stem, the more likely it is that these two will appear together in the future. This is quite evident for *anbieten* and *abriegeln*, which, despite having other possible stem-prefix pairs, make up 96.32% and 97.19%, respectively, of the total number of verb-stem and prefix pairings in the entire corpus for each stem.

Second, the total number of possible prefixes that can appear with a verb stem helps to delimit the number of possible predictions a person must make. The more separable prefixes that appear with a certain verb-stem, the more uncertainty is built into the prediction.

Third, the presence of an additional, strong co-occurrer in the sentence, adds a different layer to the prediction process. Because this was shown to be a contributing factor to successful prediction, it means that prediction ability is not simply tied to the verb-prefix combination, but rather is continuously updated as more and more information in the sentence is revealed to a person.

In addition to the overall findings, the analysis of individual items provides some further insights into the behaviors of the participants. The results show that some items, despite being relatively infrequent, were still correctly assigned to the missing target prefix (e.g. *vorlesen*, *umgehen* ), while others with similar frequencies were not (e.g. *anfügen*, *einkommen*). A closer look at the aforementioned cases provides insight into some semantic reasons behind these differences. For the verbs *vorlesen*, *umgehen*, and *einkommen*, one possibility is that the prefix has a significant impact on the semantics of the root meaning of the verb. For the first item, *lesen* [to read], becomes [to read aloud], which is involved in a slightly different context than the act of reading in and of itself. The verb *umgehen* [to be about], from the stem *gehen* [to go], takes an even more drastic shift in its core meaning when paired with *um* [about]. Thus in item 23, the probability that it could be any other preposition to express a similar meaning is seriously limited and the change in meaning is much greater than if it were to be paired with a prefix that simply expressed something like the direction of travel, which would preserve the core meaning of the verb stem. A similar change occurs with *einkommen*, which could mean “to come in” or “income” in a monetary sense. Looking at the corpus, it is actually more common to see *einkommen* as an inseparable prefix inferring a monetary interpretation than as a separable prefix which invokes a directionality to the verb *kommen* [to come]. In item 21, this strong connection to the monetary rather than directional meaning may have prodded participants against a prediction for *ein* and towards some other more frequent verb-prefix pair.

For the verb *fügen* [to attach], an interesting insight into why it was so infrequently assigned the prefix *an* [to / on] comes from the incorrect answers. Of the 47 “incorrect” responses, all 47 were *hinzu* [to there]. Looking back at the corpus, *hinzufügen* makes up 65.8% of all instances of *fügen* + separable prefix, and *fügen* as an independent verb with no separable prefix occurrs only 34,410 times in the corpus compared to the 42,206 times it appears with a separable prefix, and of those 42,206 tokens, 27,764 (65.8%) occur with *hinzu*. The connection strength between the separable prefix *hinzu* and the verb stem *fügen* is clearly driving the low accuracy on this item.

**Conclusion**

This study has provided evidence to support the claim that German speakers are able to predict sentence final separable prefixes, and that the success of prediction is mediated largely by lexical factors as they play out during sentence comprehension. The successful nature of this process argues against the low-payoff critique of previous prediction research, at least for this construction.

But the question remains whether listeners can do this actively in non-experimental settings. Because the experimental design of this program was conducted using a cloze task and it is actively asking them to predict information, they are already primed to predict a missing word, which does not mean they actively do so in everyday life in aural or visual sentence processing. However, the ability to predict others’ words could be viewed as especially important for human development and the formation of communities through alignment. As Garrod and Pickering (2014) and Pickering & Garrod (2006) explain, alignment is the basis for successful communication. This alignment occurs within multiple timescales (MacWhinney, 2014), involving phonological, syntactic, semantic, conversational, and situational processes. Successful turn-taking depends crucially on the ability to predict within less than a second the end of the previous speaker’s contribution (Levinson & Torreira, 2015). While the phonological and syntactic alignment within a conversation is important, it is through our shared life experiences that two speakers can create similar, although certainly not identical, situation models, which seems closely linked to the idea of mental models (Johnson-Laird, 1983; MacWhinney, 2005). This alignment, when successful, between two or more individuals can lead to similar mental models of a shared intentional state. This would therefore lead to a mental state that is not so different from the one presented in the visual world studies presented earlier, where the number of possible items is severely limited by the particular mental representation of the environment. If a person was to co-construct this mental representation through interaction with another person or a text, then there would be significant social and evolutionary benefits. For example, Fusaroli et al. (2012) found that linguistic alignment produced significant increases in dyad’s abilities to successfully perform tasks. This advantage is certainly in line with Tomasello et al.’s (2005) theory regarding shared intentions. Prediction can be seen as a natural result of mimicry and understanding. The greater the alignment and the more productive prediction can be, the greater ability to share intentions across individuals, culture, and time.

While there is still significant work to be done to understand whether prediction as a constantly occurring behavior, this study has provided evidence to support that long-distance lexical prediction is not only possible, but most likely a highly productive part of human communication. The finding that successful prediction of sentence final prefixes relies heavily on lexical statistics is also important, because it shows how diverse the sources of influence on a particular word are. From an emergentist, usage-based account, the dynamic nature of prediction can be linked to its usefulness in human communication, where it provides distinct support for conversational alignment.

**References**

Altmann, G. T. M. & Kamide, Y. 2007. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language,* *57*, 502–518.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes,* *28*(4), 525-560.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32.

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in cognitive sciences*, *17*(7), 348-360.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Bransford, J., Barclay, R., & Franks, J. (1972). Sentence memory: A constructive vs. interpretive approach. *Cognitive Psychology,* 3, 193-209.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Christiansen, M. H., & Chater, N. (1999). Connectionist natural language processing: The state of the art. *Cognitive science*, *23*(4), 417-437.

*Duden Online*. (n.d.) Duden.de. Retrieved 17 March 2015.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition, 10*, 101-116.

Ellis, N. C., O’Donnell, M. B., & Römer, U. (2015). Usage-Based Language Learning. In B. MacWhinney & W. O’Grady (Eds.) *The handbook of language emergence* (pp. 163-180). West Sussex, UK: John Wiley and Sons Ltd.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469-495.

Forster, K. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology 33A*, 465-495.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstroff, A., Rees, G., Frith, C., & Tylen, K. (2012) Coming to terms: Quantifying the benefits of Linguistic Coordination. *Psychological Science 23*(8), 931-9.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in cognitive sciences*, *8*(1), 8-11.

Goodman, K. (1967).  Reading: A psycholinguistic guess game.  *Journal of the Reading Specialist, May*, 126-135.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science,* *30*, 643-672.

Institut für Deutsche Sprache. (1991-2015). *COSMAS I/II (Corpus Search, Management and Analysis System)*. Retrieved From http://www.ids-mannheim.de/cosmas2/

Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kutas, M., & Hillyard, S. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11*, 539-550.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Levinson, S., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 1-17.

MacWhinney, B. (2005). Language evolution and human development. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind* (pp. 383– 410). New York: Guilford.

MacWhinney, B. (2014). Conclusions: Competition across time. In B. MacWhinney, A. Malchukov, & E. Moravcsik (Eds.), *Competing motivations in grammar and usage* (pp. 364-386). New York: Oxford University Press.

Miller, G. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2*, 217-228.

Otten, M. & Van Berkum, J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes, 45*, 464-496.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes, 11*(6), 583-588.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268* (5217), 1632-1634.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173-204.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences, 28*, 675-735.

Thurmair, M. (1991). Warten auf das Verb. Die Gedächtnisrelevanz der Verbklammer im Deutschen. *Jahrbuch Deutsch als Fremdsprache*, *17*, 174-202.

Tulving, E. & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology 66*, 319-327.

Van Petten, C. & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology 83*, 176-190.

Yoshimura, Y., & MacWhinney, B. (2007). The effect of oral repetition in L2 speech fluency: System for an experimental tool and a language tutor. *SLaTE Conference* October 1-3,pp. 25–28.

**Appendix**

Less-context item set with verbs italized, separable prefixes underlined

1. Am 1. Juli 1979 *trat* sie wieder in den Dienst der Sparkasse Rhein-Nahe ein.
2. Verschiedene Anwendungen wie das Kartenprogramm Google Maps und Googles lokale Suche *arbeiten* für Anwendungen dieser Art bei Google Earth zusammen.
3. Doch dieser Luxus *kostet* den kleinen Ort über.
4. Beim Gemüse bieten sich im Winter Rosenkohl, Feldsalat oder Chicorée an.
5. Kaum hatte sich Bohr zusammen mit zehn Teamkameraden beim 27. Silvesterlauf des TuS Waldböckelheim auf den Weg gemacht, *legte* Kommentator Willi Lange los.
6. Bei den Schweinfurter Mad Dogs *setzten* sich die Bären gestern Abend mit 5:4 (2:1, 0:2, 2:1) nach Verlängerung durch.
7. Auch hierauf spekulieren einige Firmen und *geben* nach Auftragserteilung ein vielfach teureres Nachtragsangebot ab.
8. Mit Freundin Martina Eberl, einer Profigolferin, trank er ein Gläschen Sekt und schaute lieber nach vorn: In den "heimlichen" Wünschen *kamen* auch die Olympischen Winterspiele in anderthalb Monaten vor.
9. "Auch in den nächsten Jahren wird es einen Weihnachtsmarkt vor und in der Wiedparkhalle geben", *kündigte* Salz an.
10. In einer Feierstunde *ging* Bürgermeister Karl Heinz Simon in einem Rückblick auf den beruflichen Werdegang seines Mitarbeiters ein.
11. Professor Dr. Andreas Kruse, Vorsitzender der Sachverständigenkommission der Bundesregierung, *stellte* die Ergebnisse des fünften Altenberichtes vor.
12. Die Extraausgabe in der Reihe "Blätter zum Land" *stellt* das neue Haus und die Dauerausstellung bei Hermeskeil im Hunsrück vor.
13. Auch Arbeitgeberpräsident Dieter Hundt *sprach* sich erneut für Kombilöhne aus.
14. Er studierte die Inschriften auf den Wänden des Gebäudes und *las* manche von ihnen laut vor.
15. Umtausch ist auch dort kein Problem: "In der Regel *nehmen* die Kunden direkt neue Ware mit.”
16. Die Bundesregierung *wies* Zweifel an der Verfassungsmäβigkeit zurück.
17. Demnach *riegelten* US- Soldaten die Gegend um die Abu Tajmija-Moschee im Westen Bagdads für sieben Stunden ab.
18. Kurz nach dem Start der 5000 Meter *sind* die Läufer noch dicht zusammen.
19. Aufgrund einer kommunalen Gebietsreform *schlieβen* sich die bisher selbstständigen Gemeinden Rijnsburg, Valkenbourg und Katwijk zusammen.
20. Mit einem Sack Nahrungsmittel auf dem Kopf und einer Liegematte unterm Arm *kehrt* sie in ihr Zuhause im Fischerdorf Akkrapattai zurück.
21. Und die meisten Gäste *legten* deshalb das eine oder andere Bratwürstchen unter.
22. Auβerdem *griff* das Feuer auf den Dachstuhl eines benachbarten Wohnhauses über.
23. Sie wissen die wertvollen Stücke zu schätzen und *gehen* vorsichtig mit um.
24. Das Verhältnis zu Moskau hat für Georgien auch eine innenpolitische Dimension: Mit russischer Hilfe *sagten* sich die Regionen Abchasien und Südossetien Anfang der 90er-Jahre von Tiflis los.
25. Schröder stellte dem Kreml-Herrscher indes in steter Regelmäβigkeit ein Einser-Zeugnis als Muster-Demokrat aus.
26. Wer sich bei der Dosierung unsicher ist, *greift* am besten auf anwenderfreundliche Fertigprodukte zurück.
27. Delfine müssen nämlich vieles ertragen: Die Meere sind verschmutzt, und Fischer fangen ihnen mit riesigen Netzen die Nahrung weg.
28. Und auch die Nahrungssuche *spielt* sich innerhalb der Gruppe ab.
29. Der Bundestrainer lobte “die gute Mannschaftsleistung", denn Alexander Herr und Michael Neumayer auf den Rängen 15 sowie 17 *kamen* auch im Vorderfeld ein.
30. “Dann kann ich beruhigt in Rente gehen", fügte der Werksleiter Eckhard Günner schmunzelnd an.
31. Die Reformation in Hessen im Jahre 1542 *fällt* mit dem Ende der Nonnenstiftskirche Brunnenburg zusammen.
32. Am Finaltag *greifen* dann Mannschaften wie die Oberligisten TuS Mayen und Hassia Bingen sowie Rheinlandligist FV Rübenach und der Mittelrhein-Verbandsligist VfL Rheinbach ins Turnier ein.
33. Der Vorsitzende der katholischen Deutschen Bischofskonferenz, Kardinal Karl Lehmann, *sprach* sich angesichts zunehmender Verunsicherung vieler Menschen für eine "neue Gelassenheit" aus.
34. Und die Gäste lieβen sich nicht lange bitten und *stimmten* mit ein.
35. Die Ergebnisse seiner Kontaktlinsen- und Skibrillentests *setzt* die Industrie in Spitzenprodukte um.
36. Die weit verstreute Gemeinde Elbergrund -Blickhauserhöhe *fand* sich zusammen.
37. Nach dem Motto "alles Überflüssige weg" *räumt* die neue Bundesregierung derzeit im Subventionsdschungel auf.
38. Die aktiven Musiker *stimmten* über ihren Favoriten ab.
39. Als das nicht gelang und die jungen Damen mit dem Auto wegfahren wollten, *prügelten* die Männer los.
40. Die weitaus gröβere Summe geht über Umlagen für Verbandsgemeinde und Kreis wieder weg.