Kobe University Repository : Kernel

PDF issue: 2026-11-13

Ordering of Elements in Learner Corpora

Barlow, Michael

(Citation)
Learner Corpus Studies in Asia and the World,2:127-136

(Issue Date)
2014-05-31

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81006695

(URL)
https://hdl.handle.net/20.500.14094/81006695

KOBE UNIVERSITY

Ordering of Elements in Learner Corpora

Michael BARLOW
University of Auckland

Abstract

In this paper we investigate the use of adverbs by L2 writers, using data from French, Polish and Swedish speakers taken from ICLE. In addition, we make use of data from LOCNESS as a comparative NS corpus. Rather than calculate overuse or underuse for the writing as a whole, we provide a more specific analysis and relate frequency of adverb use to position in the sentence: 1st word, 2nd word, etc. This provides a more nuanced view of differences among L2 writers and also links their writing to text structure, which in this case is at the level of the sentence. This also has the potential to provide some insights into the rhetorical flow of writing among the different sub-corpora.

Keywords

Adverbs, ICLE, Sequence, Ordering, Sentence structure

I Introduction

There is a substantial literature describing the characteristics of L2 writing with respect to the overuse or underuse of particular adverbs (Hinkel 2002). This line of research has been made possible by the availability of learner corpora such as ICLE. In this paper we further examine L2 writing with respect to adverb usage, paying particular attention to the interaction between the choice of adverb and position within sentences. Conceptually, the notion of overuse or underuse is quite straightforward. It is clear that L2 writers have a tendency to favour certain words or phrases and avoid others. Some writing L2 patterns are difficult to pick out and a corpus-based analysis has been very useful in marshalling quantitative data to provide an overview of patterns of use. By examining data on the positioning of adverbs within the sentence, we can obtain further insights into the patterns used by L2 learners. The investigation of adverbs also serves to illustrate a method of learner corpus analysis that reinstates data that is typically backgrounded: information about text structure and textual order.

II Literature Review

There have been some L2 writing studies on the ordering of words – adverbs, subordinators and other connectors -- in texts (Aarts & Granger 1998, Altenberg & Tapper 1998) and on the use of explicit markers of the logical or discourse relations within a text (Ädel 2006, Flowerdew 2008). Research on the ordering of words has typically focussed on sentence-initial and sentence-final positions, as these are relatively easy to identify.

The literature on adverb use by L2 learners is quite extensive, as noted above, and given the nature of adverbs and adverbials, it is not surprising that the topics covered are quite varied (Altenberg and Tapper 1998, Chen 2006, Crawford 2008, Hinkel 2003, Narita & Sugiura 2006, Pérez-Paredes & Díez-Bedmar 2012, Tankó 2004).

III Research Design

The focus for the present study is L2 writing and for this we analyse samples of argumentative writing in English by French, Polish and Swedish speakers taken from the International Corpus of Learner English (Granger et al. 2002), along with some equivalent essays written by native speakers taken from the Locness Corpus. (See Table 1.) The advantage of using this data is that the type of writing is the same across the different subcorpora, as is the proficiency level of the second language learners. Thus we can carry out a comparative analysis. We can also contrast the patterns of usage of the different groups of learners and make comparisons with the English native speakers (NS).

Table 1 Corpus data

L1	No. essays	No. words	No. sentences	No. words per sentence
French	382	221525	11411	19.4
Polish	363	235543	12154	19.4
Swedish	285	165863	8223	20.1
NS	318	220894	11193	19.7

The essays were annotated using the CLAWS7 POS tagset and the writing was then analysed using WordSkew, which allows the user to determine the frequency of words, phrases or POS tags across portions of different textual units: sentences, paragraphs, or the complete text. The "portions" can either be calculated in relation to equal division of the unit --- first 10%, second 10%, etc --- or as absolute positions such as first word in the sentence, first sentence in the paragraph etc. Thus it is possible to search for positions "1 2 Mid -1 #" where # stands for last position and -1 is the penultimate

position. The software gives the results as histograms (and tables). Clicking on a particular bar of the histogram reveals the concordance lines for that position. In this study a search for the position of a linguistic expression within a sentence will process around 10,000 sentences for each L1. Since the student sentences tend to be quite long, we can focus on a range of positions in the sentence: 1 2 3 4 Mid -3 -2 -1 #. The label "Mid" differs from the others and covers several words: those that are in any of the other specific positions. In order to make "Mid" comparable with the other single word slots when calculating the results, the frequency of usage for the Mid position is based on a per word calculation. Thus, to simplify, we can say that if there is one adverb occurring in the ten words comprising Mid, then the frequency would be 0.1.

This method, which combines word counts and sentence counts as well as tracking position information, provides a finer-grained approach to overuse and underuse and links the usage with position in the sentence. Naturally, it is possible to calculate the relative frequency of use of a word or word class in terms of words per 1000 Sentences. What we do, however, is calculate the frequency per 1000 sentences for each position in the sentence.

IV Results and Discussions

The constraints on ordering of elements within a sentence complex; they are partly grammatical and partly discourse- or genre-related and these choices and constraints are a part of what must be learned in order to write in a second language. We can present one illustration here by examining the top-ranking two-word adverbs in positions 1 2 3 4 Mid -3 -2 -1 in the sentences. (Sentences with fewer than 8 words are discarded for the purpose of making these calculations.). For NS, we obtain the following sequence for these positions, counting only rank number 1: "*of course, of course, no longer, no longer, at least, at least, no longer, as well.*" The choice of adverbs and their order clearly shows the connection between the position of the adverbs and the typical ordering of discourse in the presentation of the writers' opinions. We can contrast this with the most frequent two-word adverbs used by the French speakers: "*of course, of course, of course, of course, at least, at least, at least, at all.*" We see that these learners of English are following the same general pattern, but with a more limited set of forms, reflecting a less elaborated rhetorical structure.

The graph in Figure 1 shows the distribution of simple adverbs in sentences. What is striking is that the general configuration is remarkably similar for the different groups and follows a max-low-high-falling trajectory. This similarity suggests that all the L2 writers have mastered the basics of English sentence organisation and that differences relate to preferences, presumably related to L1 or instruction, rather than a lack of knowledge of the grammar of sentences. For all groups of writers, position 1 is

associated with the most frequent use of adverbs. Compared with a general count of adverb use by total words or total sentences, the positional analysis provides a clearer picture of overuse and underuse and here we can see that the distinctions among the different groups is most pronounced in positions, 1, 3, 4, with position 1 being the most markedly divergent. The Polish speakers have the greatest preference for sentence-initial adverbs, followed by the French speakers. In this case, the Swedish and NS writing is indistinguishable. It appears that French and Speakers have a greater preference for adverbs in the third word slot in the sentence.

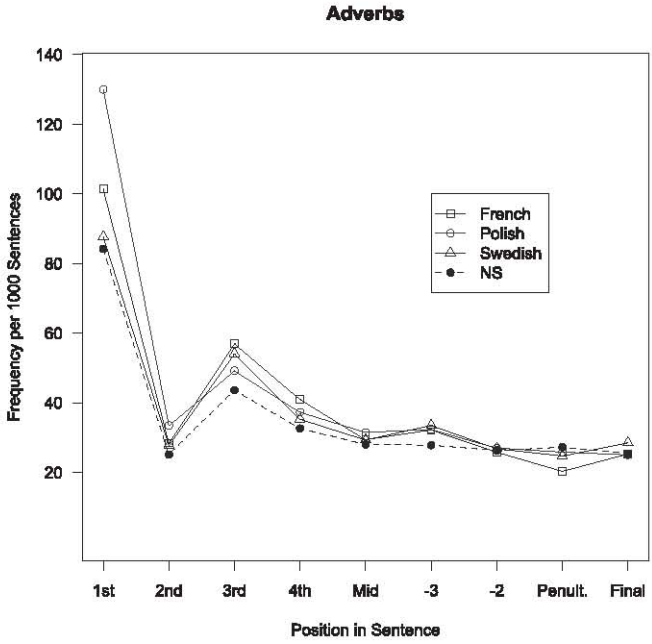


Fig. 1 Distribution of adverbs in sentences

It is important to remember that the frequencies in Figure 1 are an average over all the essays in each sub-corpus. Figure 2 provides a view of the amount of individual variation. To show this variation, the usage of adverbs in sentences is calculated for each of the 363 essays and the results represented as box-plots. The horizontal bar in the box shows the median value. The bottom of the box represents the first quartile; the top represents the third quartile. Thus the height of the box gives some idea of the spread of frequency values. The small circles represent outliers. If we focus on the

median values, we can see that the variation in frequency through the sentence is less marked than when observing averages (in Figure 1). We can also see that in position 1, there are a number of outliers. That means that a few student writers are using a very high number of sentence-initial adverbs and this is "inflating" the average. It would be better to use median rather than average, but for computational convenience we will continue with the use of averages. The pattern of variation of Figure 1 and Figure 2 is similar enough for our purposes.

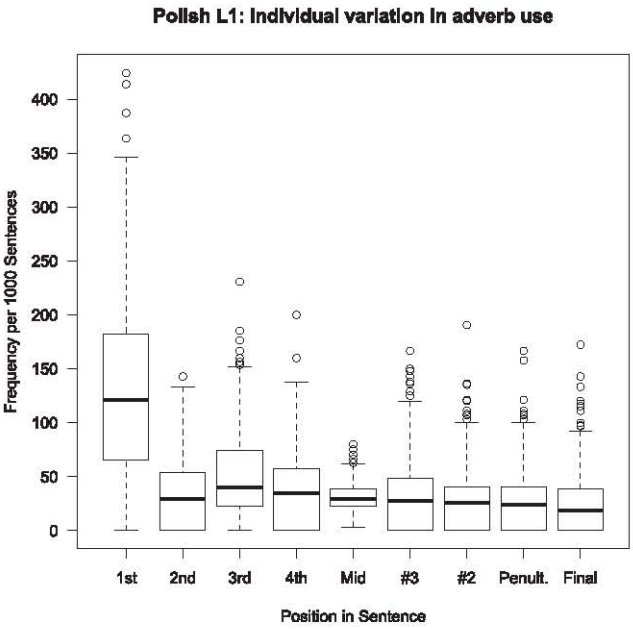


Fig. 2 Boxplot of adverb use by individual Polish speakers

Figure 3 shows the distribution of time adverbs. The trajectory is broadly similar to that of adverbs in Figure 1 but with a marked increase in sentence-final position. The frequencies are low, but it is clear that the L2 students tend to overuse time adverbs in sentence-initial position. The top ranked sentence-initial time adverbs are *nowadays, now, today* (French L1); *nowadays, then, now* (Polish L1); *today, then, now* (Swedish L1); and *now, then, today* (NS). In final position, the top-ranked time adverbs are *again, before, nowadays* (French L1); *nowadays, before, again* (Polish L1); *today, again, before* (Swedish L1); and *today, again, now* (NS). Apart from the L2 writers' preference for *nowadays*, the different groups tend to use the same set of adverbs.

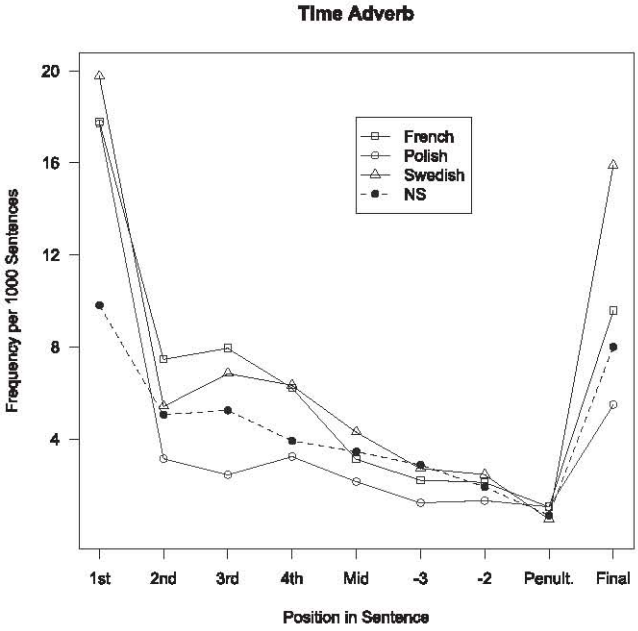


Fig. 3 Distribution of time adverbs in sentences

Apart from the peak due to modification of the main verb that occurs with both time and degree adverbs, the distribution of the two types shown in Figure 4 tend to complement each other, reflecting the differences in their main function in the sentence. (Degree adverbs are *so too very*, etc.) As with many of the distributions, the trajectory of use by NS is typically lower and flatter than that of the L2 writers.

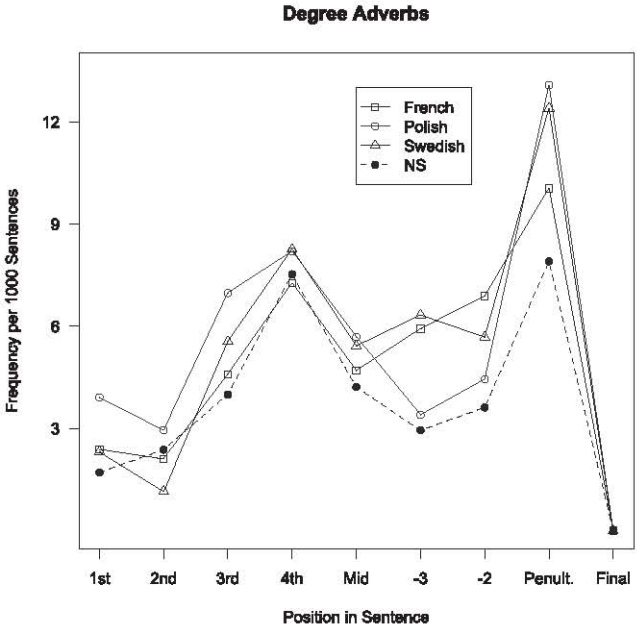


Fig. 4 Distribution of degree adverbs in sentences

Figure 5 presents the distribution of two-word adverbs such as *of course* and *as well*. We referred to a pattern of use above and the graph in Figure 5 shows the overall distribution of the adverbs. We see a general overuse by L2 writers, but also a difference in patterning. In sentence-initial position, the French writers show the most frequent use of two-word adverbs, while in sentence-final position the usage is similar to that of NS. On the other hand, Polish and Swedish L2 writers use the adverbs at about the same rate as NS in position 1, but have a marked overuse in sentence-final position.

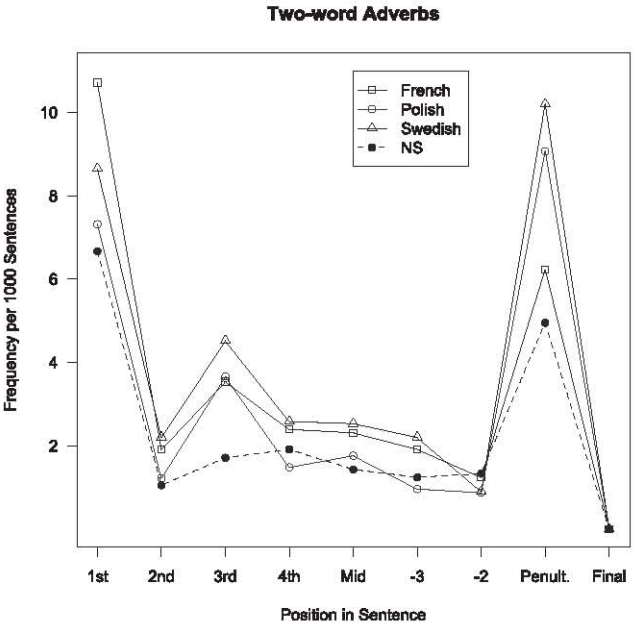


Fig. 5 Distribution of two-word adverbs in sentences

Table 2 shows the most frequent two-word adverbs used by the different writers in sentence-final position and we can see that the choice of adverb is broadly similar with *at all* and *as well* featuring frequently. We can observe the use of *of course* by the Swedish speakers, which is somewhat unusual in this position.

Table 2 Two-word adverbs in sentence-final position

French Speakers		Polish Speakers		Swedish Speakers		NS	
Percent	Adverb	Percent	Adverb	Percent	Adverb	Percent	Adverb
28.7%	*at all*	35.5%	*as well*	30.0%	*as well*	37.2%	*as well*
21.2%	*in general*	26.9%	*at all*	25.0%	*at all*	25.4%	*at all*
9.1%	*as well*	7.6%	*any longer*	8.7%	*of course*	9.8%	*vice versa*

Plotting the frequency of use of *of course* (Figure 6) reveals a general overuse by French and Swedish speakers, especially in sentence-initial position. The small increase in use of the adverb by Swedish writers in sentence-final position contrasts with the zero use by NS.

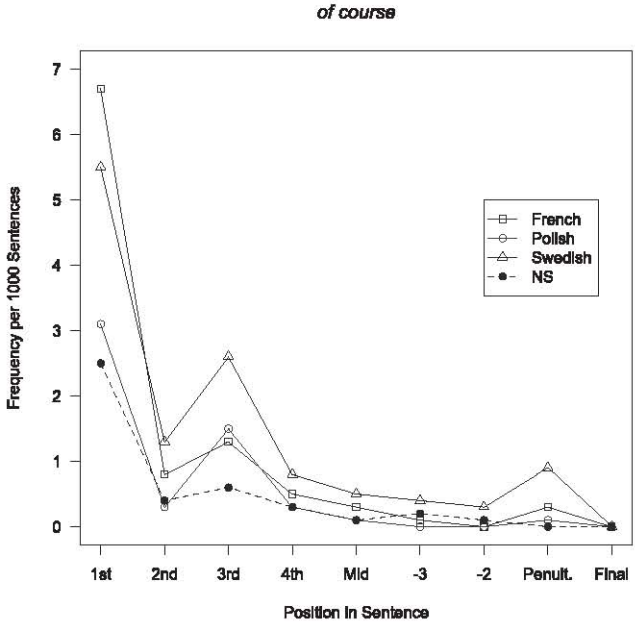


Fig. 6 Distribution of *of course* in sentences

V Conclusion

We have explored some differences in the use of different types of adverbs by writers with a French, Polish, Swedish and NS background. The main thrust of this paper, however, is to show how the addition of position information (with respect to sentences) adds to our knowledge of L2 writing and provides a further way to exploit learner corpora.

References

Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132–141). London & New York: Addison Wesley Longman.

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam & Philadelphia: John Benjamins.

Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learner's written English. In S. Granger (Ed.), *Learner English on computer* (pp.

80–93). London & New York: Addison Wesley Longman.

Chen, C. W.-Y. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics, 11*(1), 113–130.

Crawford, W. (2008). Place and time adverbials in native and non-native English student writing. In A. Ädel & R. Reppen (Eds.), *Corpora and discourse: The challenges of different settings* (pp. 267–289). Amsterdam & Philadelphia: John Benjamins.

Granger, S., Dagneaux, E., & Meunier, F. (Eds.) (2002). *International corpus of learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Flowerdew, L. (2008). *Corpus-based analyses of the problem-solution pattern: A phraseological approach*. Amsterdam & Philadelphia: John Benjamins.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ.: Lawrence Erlbaum Associates.

Hinkel, E. (2003). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics, 35*(7), 1049–1068.

Narita, M., & Sugiura, M. (2006). The use of adverbial connectors in argumentative essays by Japanese EFL college students. *English Corpus Studies, 13*, 23–42.

Pérez-Paredes, P., & Díez-Bedmar, M. B. (2012). The use of intensifying adverbs in learner writing. In Y. Tono, Y. Kawaguchi, & M. Minegishi (Eds.), *Developmental and crosslinguistic perspectives in learner corpus research*. (pp. 105–123). Amsterdam & Philadelphia: John Benjamins.

Tankó, G. (2004). The use of adverbial connectors in Hungarian university students' argumentative essays. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 157–181). Amsterdam & Philadelphia: John Benjamins.