

Using Corpora to Analyze Learner English: The Case of the Dative Alternation

Giovanna Silvestro

Traditionally, contrastive analyses of learner language and native language have been conducted relying solely on intuition tests. Adopting corpus linguistics methodologies, this paper attempts to investigate how French learners of English acquire the dative alternation. Two L2 corpora of spoken English have been analyzed and, then, compared to a corpus of native English. An examination of their concordance lines and a comparison of their frequencies in both learners' and native speakers' conversations have revealed that the distribution of the to-variant and the double object variant in learner language might depend on several factors: a transfer from the learners' L1, an awareness of the native speakers' preferences, a low proficiency level or a personal stylistic choice. In spite of its limitations and the lack of a straightforward interpretation of the findings, this research can offer an insight into the learners' actual use of the target language and can constitute a starting point for further studies.

Keywords: Second language acquisition; Language transfer; Corpus Linguistics

Over the past thirty years, corpus linguistics has shown an increasing interest in learner language. Granger (2004) and Gilquin and Granger (2015) have shown to what extent Computer Learner Corpus (CLC) research, or Learner Corpus Research (LCR), can be referred to as "a new way of thinking about learner language" (Granger 2004: 123). They have highlighted the features differentiating CLC or LCR from the traditional research methods of Second Language Acquisition or Foreign Language Teaching, whose purpose is a description of the learners' linguistic competence, and they have advocated for the advantages of using computers in analyzing learner language.

Part of the aim of this project is to focus on French learners of English and to examine, adopting corpus linguistics techniques, how they deal with the English dative alternation. A considerable amount of literature has been published on the alternating morpho-syntactic realization of the goal or recipient of an action in English. Scholars have analyzed the properties of this alternation in many different ways in order to point out the semantic, lexical and pragmatic restrictions ruling it (see Krifka 2004). In their major study about dative alternation, Levin and Rappaport Hovav's (2008) have argued for a verb-sensitive analysis of the phenomenon, in particular, starting from Jackendoff's (1990) approach, they propose a multiple meaning approach to the dative alternation, according to which the two variants in the expression of same argument in

dative constructions are associated with two different meanings unequally distributed. They claim that the meaning of the dative verb itself determines the distribution of the double object variant (DOC), always bearing a meaning of caused possession, and the to-variant, which can bear both a caused possession or a caused motion meaning. Along with a large volume of studies describing the dative alternation in English, a wide range of articles and books focusing on the reasons why English and other languages (particularly Germanic) present two distinct options for expressing the recipient in dative constructions, while other languages, especially Roman languages, do not is available (Harley 2003; Haspelmath 2005; Siewierska 1998). Levin and Rappaport Hovav (2008) and other researchers (McFadden 2002; Polo 2002) have attempted to investigate the factors determining the rise of the two variants in the linguistic history of English and their distribution, underlining the key role played by the interaction of semantic connectedness, heaviness and information structure, however little is known about the reasons for the lack of such alternation in languages like French.¹

Assuming this asymmetry between English and French, this paper sets out to provide an insight into how intermediate-advanced French learners of English use the two variants available in English to express the indirect object in dative constructions. Whilst few studies about the acquisition of the dative alternation have been conducted in the past, they have been exclusively descriptive in nature and they have not been systematically based on large collections of speech produced by foreign language learners. In her research, Marzukewich (1984) has suggested that the learnability of the alternation should be studied within the theoretical framework of markedness. Drawing on the support of developmental studies, she maintains that unmarked structures in the Core Grammar of a language, such as the to-variant, are learned before marked structures, like the DOC-variant, an order followed both by native speakers and second language learners. Commenting on Marzukewich's (1984) results, Kellerman (1985) has argued that the undeniable flaws in her analysis make her conclusions unacceptable. In his review, he lists three weak points of Marzukewic's (1984) article: firstly, the lack of an adequate number of examples showing how the lexical constraints governing the alternation operate;2 secondly, the use of tests based on intuitive judgments about the grammatical acceptability of preconstructed sentences (see Kellerman 1984); and thirdly, the lack of an objective statistical treatment of the data and the exclusion of any possible influence of the L1 in the process of learning.

In view of all that has been mentioned so far, it seems clear that there have been no controlled studies which compare differences in the use of the two variants of the dative alternation in native speaker language and learner language. Therefore, the main purpose of this research is to develop an understanding of the way in which learners of English handle this alternate argument realization and to compare their linguistic choices with English native speakers' linguistic behavior. In order to fulfill these objectives, corpus linguistics methodologies have been adopted to analyze in a contrastive way two corpora of spoken learner English and a corpus of spoken native English. Due to practical constraints, this paper cannot provide a parallel analysis of written native and learner English, nevertheless, it can offer an interesting contribution to the growing area of research on learner corpora.

¹ Note that Old French showed the dative alternation, while present-day French lacks it. Different hypotheses about this linguistic change have been advanced (see Kayne 1984; Tremblay 1991).

² Marzukewich (1984) holds the view that dative alternation is better explained in terms of a lexical approach (Oherle 1976).

Data and Methodology

Corpus selection

Data for this research were gathered from multiple sources. Three existing corpora were used: the French Learner Language Oral (FLLOC) corpora,3 the Backbone corpus4 and the spoken component of the British National Corpus (BNC).⁵ The FLLOC is a growing web-based freely available database of learner corpora among which only the University of East Anglia (UEA) corpus was selected for the current investigation. The UEA corpus contains recordings of oppositional talks held by learners of French and English collected between November 2002 and December 2004 in the School of Language, of Linguistics and Translation Studies at the University of East Anglia. For the purpose of analysis, exclusively transcribed data from L2 speakers of English, whose mother tongue was French, were considered. The Backbone corpus is a free online corpus of video-recorded interviews with a specific pedagogical design. It comprises data from native speakers of various European languages and English as Lingua Franca (ELF) speakers. Thanks to the "section search" mode, one of the online Backbone search tools, it was possible to display only results from native French ELF speakers and to access to the transcriptions of the multimedia files. The BNC is a monolingual general corpus which covers British English of the late 20th century. Its spoken section consists of approximatively 20 million words, i.e. 10% of the whole corpus, and it comprises transcriptions of formal and informal language. In this study, only one file, KBO, was selected.

Representation	Name of the corpus	Number of word tokens
Learner English	UEA	5793
Learner English	Backbone	8793
Native English	BNC (KB0)	44076

Table 1. Number of word tokens in the corpora analyzed.

Corpora were selected on the basis of a degree of homogeneity of the linguistic competence of the learner speakers (an intermediate-advanced level was preferred), while no particular attention was paid to the genre they represent since genres seem not to have a role in the distribution of the two variants of the alternation in native English. Additionally, the possibility of freely accessing them also contributed to the choice of the three electronic collections of spoken language cited. As Granger (2004: 129) underlines, CLC researchers have to face the paradox "that while there is an abundance of learner corpora, hardly any of it is available for academic research". In the next section, the methodologies applied to investigate the selected corpora, which could be referred to as comparable corpora, will be further clarified.

Methodology

Traditionally (see Marzukewich 1984), the acquisition of the dative alternation by the French-speaking learners of English has been assessed through the use of tests based on intuition. However, there are certain drawbacks associated with that type of analysis, noticeably the fact that

³ The French Learner Language Oral Corpora are available online at http://www.flloc.soton.ac.uk/ (Last accessed: May 20, 2017). Further information can be also found on this website.

⁴ The Backbone corpus is accessible online at http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp. (Last accessed: May 20, 2017). Backbone SearchTool is also freely available on the website.

⁵ The spoken component of the British National Corpus is accessible online at http://bncweb.lancs.ac.uk (Last accessed: May 20, 2017).

its outcome does not reflect the actual use of language by learners, but it merely shows instances of their intuitions about the L2. Thus, for this project corpus linguistics methodologies, involving both a quantitative and a qualitative analysis of the data, were used.

After collection, learner corpora were cleaned using TextPad8 (Helios Software Solutions 2017), a text editor for Windows computers comprising a tool used to replace, or to delete, regularly occurring expressions in large portions of texts.⁶ Metadata providing information about the speakers' name, age, nationality, job and the date of the recording were removed. Furthermore, other regular expressions in the corpora, for example, introductory paragraphs about the topic of the speech, were found and eliminated, too. Although data from the three corpora were stored in three separated files, in their analysis they were divided into two parts: one representing learner language (the UEA corpus and the Backbone corpus) and the other representing native English functioning as a reference corpus (the spoken component of the BNC). A text retrieval software, AntConc (Anthony 2014), was used to carry out the examination of the dative constructions in the corpora. A dative construction implies the presence of a semantic trivalent dative verb showing a syntactic valency of 3, too. This means that the head verb is accompanied by three syntactically expressed arguments: a subject, a primary object and a secondary object. As in English with certain dative verbs, the secondary object can be expressed both in the form of an indirect or direct object and the aim of this research is to show how learners deal with this alternation, it seemed necessary to identify a list of dative verbs allowing the dative alternation. Thus, Levin and Rappaport Hovav's (2008: 134) classification was chosen and used to derive the search terms to be used in the software.

Give-type	Send-type	Throw-type
Verbs of giving: give, hand, lend, loan, pass, rent, sell.	Verbs of sending: forward, mail, send, ship.	Verbs of causation of motion: fling, flip, kick, lob, slap, shoot, throw, toss.
Verbs of future having: allocate, allow, bequeath, grant, offer, owe, promise.	Verbs of instrument of communication: e-mail, fax, radio, wire, telegraph, telephone.	Verbs specifying the direction of caused motion: <i>bring</i> , <i>take</i> .
Verbs of communication: tell, show, ask, teach, read, grant, offer, owe, promise.		

Table 2. A classification of dative verbs allowing the dative alternation.

To some extent, this may appear an incomplete account, however, for the purpose of this paper, it was considered to be a good point to start from. To ensure comparability of the data, a

⁶ TextPad is available online at https://www.textpad.com. A trial version can be freely downloaded.

preliminary research was conducted in both the learner and the native speaker corpora so as to find which dative verbs occurred in them. Using the Concordance tool in AntConc, each verb in the table was searched in all its forms thanks to the help of a wildcard. The analysis proceeded then with a close examination of the concordance lines in order to avoid the cases in which dative verbs were not used in dative constructions. Only the dative verbs occurring in both corpora were taken into consideration, as this would have helped to compare the two corpora.

Granger (2004) and Gilquin and Granger (2015) identify several advantages of comparing learner corpora to normative references corpora. In particular, this contrastive approach is defined Contrastive Interlanguage Analysis (CIA) and it includes not only the comparison between different varieties of interlanguages but also between learner language and native speaker language. For this paper, it was decided that the latter one was the best method to answer its research questions. One of the major weaknesses of CIA, however, is the fact that it overlooks the impact of the transfer from the L1 in the learning process. Thus, a more rigorous method, the Integrated Contrastive Model has been introduced to allow the researchers to consider both the nature of the learners' L1 and their L2 (see Gilquin and Granger 2015; Gilquin 2008). This paper attempts to explore its research questions adopting such a perspective, though only to some extent, because the interactions between the learners' first language and their target language should be further investigated.

Comparisons between the two groups of data were made on the basis of the frequency values of the two possible variants in the expression of a secondary object in dative constructions. Since the two sets of data did not have the same size, the need for reliable and comparable results required a normalization of the raw frequencies of each variant for each dative verb. The obtained ratings were then used to compare learners' use of English and native speakers' behavior with respect to the dative alternation. As previously noted, the analysis was limited to spoken language which, as Granger (2004: 125) notices, arises a series of problem specifically for its variability:

Learner language is highly variable. It is influenced by a wide variety of linguistic, situational and psycholinguistic factors, and failure to control these factors greatly limits the reliability of finding in learner language research. [...] While many of these variables are also relevant for learner corpus building, the specific nature of learner language calls for the incorporation of L2-specific variables [...] pertaining the to the learner or the task.

However, the overwhelming majority of learner corpora are not designed with a specific research orientation, conversely, they show a more generic nature, and this could affect in some ways the findings resulting from them. And yet, it is a matter of fact that a learner corpus is a specialized one, since it is representative of a particular variety of language, hence it cannot be ignored that it has all the disadvantages but also all the advantages of such a type of corpora.

Analysis of the data

Corpora were first analyzed to determine the distribution of trivalent verbs allowing the dative alternation in both learners' and native speakers' language use. In the learner corpora (UEA, Backbone) 13 dative verbs of the list proposed in Levin and Rappaport Hovav's (2008: 134) classification occurred (*give, sell, offer, send, throw, bring, take, allow, tell, show, ask, read, write*). However, only 5 of them occurred in dative constructions (both variants of the dative alternation were considered at this point of the analysis). These 5 verbs (*give, sell, send, offer, bring*) became the search terms for the examination of the BNC corpus, where only 4 of them occurred in dative constructions. In table 3, the number of occurrences of these verbs in both sets of data is summarized. However, the results obtained in this preliminary analysis show only the raw frequencies of the dative verbs selected and of their occurrences in dative constructions, thus making comparisons on

⁷ Frequencies were normalised per a desired size of 10,000 words.

this basis would be a too approximative type of approach. Additionally, these numbers do not suggest any answer to the research questions of the present study, but they constitute the basis of a further investigation.

Corpora	UEA - Backbone		BNC	
Dative verb	Number of occurrences	Number of dative constructions	Number of occurrences	Number of dative constructions
GIVE	6	4	42	19
SEND	1	1	4	2
OFFER	1	1	15	3
BRING	4	4	29	11
SELL	2	1	2	0

Table 3. Occurrences of dative verbs in the learner corpora and the native speaker corpus.

The first set of analysis was a combined analysis of the two learner corpora which was carried out to distinguish the cases in which the to-variant was preferred from the ones in which the DOC-variant was used.

Dative verb	Number of dative constructions	Number of TO- variant	Number of DOC- variant
GIVE	4	2	2
SEND	1	1	0
OFFER	1	1	0
BRING	4	4	0
SELL	1	1	0

Table 4. Distribution of the variants in the learner corpora.

In Table 4 the distribution of the two variants of the dative alternation is shown for each verb. What is interesting about it is the case of *give*. The example in (1) illustrates the concordance line of an occurrence of *give* used in the DOC-variant, while that in (2) shows a case in which *give* is used in the to-variant. Both of them are taken from the Backbone corpus.

- (1) It's the best way to give my child something.
- (2) Because people give their life on the web to everyone.

Learners are expected to transfer unconsciously their linguistic habits into the target language (see Lado 1957; Selinker 1966; Kellerman 1979), thus French speakers who learn English are expected to overlook the dative alternation just because of the lack of such alternation in their native language. Nevertheless, learners in this corpus seem to be aware of its existence and they show no preference in the selection of one of the two variants. French has a unique possible structure to express the receiver or the goal of an action, i.e. the prepositional phrase whose head is the preposition \dot{a} , and, as Marzukewich (1984) has argued, the equivalent prepositional phrase in English (to-variant) is expected to be preferred by French learners of English. Interestingly, while data from the learner corpora related to verbs like *sent*, *offer*, *bring* and *sell* conform to these expectations, the findings about the verb *give* suggest that that is not always the case.

Frequencies in Table 4 could be compared to those in Table 5 representing the frequencies of the variants in the native speakers' use of the language. However, even if a comparison of the two sets of raw frequencies may be done on a purely intuitive basis, it cannot provide statistically relevant information, due to the different size of the corpora.

Dative verb	Number of dative constructions	Number of TO- variant	Number of DOC- variant
GIVE	19	6	13
SEND	2	2	/
OFFER	3	1	2
BRING	11	8	3

Table 5. Distribution of the variants in the native English corpus.

As pointed out in the introduction, one of the aims of this paper is not only to examine how learners deal with the dative alternation, but also how they deal with it if compared to native speakers. The next part of the analysis, then, seeks to reveal what type of correlation exists between the languages of the two groups of speakers. When comparing two or more corpora different in size, it is necessary to normalize raw frequencies. Only when normalized frequencies are calculated, it is possible to proceed with a contrastive analysis. Table 6 offers a summary of the data obtained from the normalization of the frequencies:

DATIVE VERB		NATIVE SPEAKERS	FRENCH LEARNERS
CIVIE	TO-variant	1.36	1.37
GIVE	DO-variant	2.94	1.37
SEND	TO-variant	0.45	0.68
	DO-variant	0	0
OFFER	TO-variant	0.22	0.68
	DO-variant	0.45	0
BRING	TO-variant	1.81	2.75
	DO-variant	0.68	0

Table 6. Normalized frequencies of the two variants of the dative alternation in learner English and native English.

With regard to the verb *give*, the to-variant seems to be used approximatively with the same frequency by learners and native speakers of English. As it was previously said, what is remarkable is the balance in the distribution of the two variants in the learner corpora, which differs, however, from the preference for the DOC-variant recorded in the language of the native speakers.

Other interesting observations can emerge from the data concerning the verb *send*. Despite a minor discrepancy in the values of the frequencies, both learners and native speakers show a strong preference for the to-variant. In a corpus-based study, Wasow (2002) attempts to offer an overview of the verb-specific preferences for a certain variant in English, more specifically, according to him, a verb like *send* tends to favor the to-variant in English because of the information structure generally associated with it. In this sense, it could be argued that learners have gathered, consciously or unconsciously, the natives' habit to use a prepositional phrase instead of a double object with the verb *send*. However, attracting, these results need to be examined with caution, noticeably because the possible interference of the L1 in the use of English L2 cannot be completely ruled out. Furthermore, this hypothesis cannot be confirmed by the relatively small corpora analyzed, a wider investigation on this issue is therefore recommended.

The case of *offer* suggests that the transfer from the L1 to the L2 actually plays a role in the use of the target language. The exclusive choice of a prepositional phrase to express the secondary object of a dative construction may be read as a lack of proficiency. Indeed, native speakers of English use the double object variant two times more frequently than the to-variant and this tendency is

completed inverted in the learners use of the same verb. In such a case, it is difficult to postulate a clear-cut interpretation of the findings. They can be affected by the L1 of the learners, or they can derive from individual stylistic choices and in LCR "the reliability of the results is not guaranteed if possible differences between individual learners are disregarded" (Gilquin and Granger 2015: 429).

Bring poses an equally interesting, yet different, problem. In the BNC *bring* is three times more often associated with a prepositional phrase headed by *to* than with a double object construction. However, some instances of the verb used in the DOC-variant are recorded.

On the contrary, in the learner corpora, there is a high frequency of the to-variant, even higher than the BNC's one, which contrasts with the total absence of cases of *bring* used in the DOC-variant. There are two likely causes for the lack of any double object construction to express the secondary object of the verb *bring*: the first one is linked to the learners' L1 and its interference as well as their personal stylistic choices; the second one has to do with the learners' level of proficiency which may be low to some extent, so that L2 speakers of English cannot gather such unconscious tendencies in the target language.

Again, caution must be applied when an attempt is made to interpret the findings derived from such a limited collection of data. More profound and detailed future analyses are therefore required.

Conclusion

The present study was designed to examine how intermediate-advanced French learners of English deal with the two variants of the English dative alternation and to compare their choices to the native speakers' ones, so as to understand whether they use the alternation like the native speakers' do or they simply reproduce the structures of their mother tongue. The corpus-based analysis conducted has revealed that learners not only seem to be aware of the existence of a closed class of dative verbs allowing the dative alternation, but also of the fact that some of them prefer to occur in a variant more than in the other one. However, this awareness emerges particularly from the data concerning the verbs *give* and *send* and seems to lack in the use of verbs like *offer* and *bring*. The reason for this inconsistency may lie in the casualness of the distribution of the two variants in learner language or in the expression of personal stylistic preferences by the learners. In fact, the study of this kind of phenomenon should start from the understanding that linguistic habits or preferences do not express the norm, but they show the actual use of a certain language by native speakers. Therefore, this paper does not intend to express value judgments on learner language, on the contrary, it aims to look at and to describe learners' use of language through an objective lens, that is through corpus linguistic methodologies. As it was explained in the previous section, the findings from this study cannot provide straightforward answers to the research questions posed at the beginning, yet they can offer an interesting contribution to the discussion about the acquisition of argument alternations.

It must be acknowledged that the generalisability of these results is subject to certain limitations. To start with, the study is based on a limited number of data because of technical difficulties in accessing bigger learner corpora. In addition to this, the selected corpora comprise only spoken language, that is subject to high variability if compared to written language. Several disparate factors can influence the utterance of speech, while, due to its own nature, written language can be more controlled at least to some extent. Thus, a further study investigating the use of the dative alternation in written texts would be interesting and perhaps revealing. Another weakness of this research it that it was not specifically designed to evaluate the impact of extra-linguistic factors, an issue which would be worth investigating in order to account for the vague interpretations given to the data analyzed.

This research has thrown up many other questions in need of further investigation instead of answering its own. Nevertheless, it seems to have accomplished the purpose to offer some insight into the actual way learners handle the dative alternation relying on empirical data, thus

representing a step forward with respect to the previous studies on the issue, mainly based on intuition-based approaches.

References

- Anthony, L. 2014. *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/.
- Gilquin, G. 2008. Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In M.B. Dez-Bedmar, G. Gilquin and S. Papp (Eds), *Linking Up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi, 3–33.
- Gilquin, G. e Granger, S. 2015. Learner language. In D. Biber and R. Reppen (Eds), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 418–435.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar approach to argument structure.* Chicago, IL: University of Chicago Press.
- Granger, S. 2004. Computer learner corpus research: current state and future prospects. In U. Connor and T. Upton (Eds), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 123–145. DOI: 10.1163/9789004333772_008
- Harley, H. 2003. Possession and the double object construction. In P. Pica and J. Rooryck (Eds), *Linguistic Variation Yearbook* 2. Amsterdam: John Benjamins, 31–70.
- Haspelmath, M. 2005. Ditransitive constructions: The verb "give". In M. Haspelmath, M. S. Dryer, D. Gil and B. Comrie (Eds), *Words atlas of language structures*, Oxford: Oxford University Press, 426–429.
- Helios Software Solutions 2017. *Text Pad* (version 8.1.2) [Computer Software]. Preston, Lancashire: Helios Software Solutions. Available from https://www.textpad.com.
- Jackendoff, R. S. 1990. Semantic structures. Cambridge, MA: MIT Press.
- Kayne, R. 1984. Connectedness and Binary Branching. Dordrecht: Foris.
- Kellerman, E. 1979. Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition* 2(1): 37–57.
- Kellerman, E. 1984. The empirical evidence for the role of the L1 in interlanguage. In A. Davies, C. Criper and A. Howatt (Eds), *Interlanguage: The State of Art*. Edinburgh: Edinburgh University Press, 98–122.
- Kellerman, E. 1985. Dative alternation and the analysis of data: a reply to Marzukewich. *Language learning* 35(1): 91–101.
- Krifka, M. 2004. Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics* 4, 1–32.
- Levin, B. and Rappaport Hovav, M. 2008. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 44: 129–167.
- Marzukewich, I. 1984. The acquisition of the dative alternation by second language learners and linguistic theory. *Language Learning* 34: 91–109.
- McFadden, T. 2002. The rise of the *to*-dative in Middle English. In D. Lightfoot (Ed.), *Syntactic effects of morphological change*. Oxford: Oxford University Press, 107–123.
- Oherle, R. T. 1976. The grammatical status of the English dative alternation. Ph.D. dissertation, MIT.
- Polo, C. 2002. Double objects and morphological triggers for syntactic case. In D. Lightfoot (Ed.), *Syntactic effects of morphological change*. Oxford: Oxford University Press, 124–142.
- Siewierska, A. 1998. Languages with and without objects: The Functional Grammar approach. *Languages in Contrast* 1: 173–190.
- Tremblay, M. 1991. Alternances d'arguments internes en français et en anglais. *Revue Québécoise de Linguistique* 20: 39–53.
- Wasow, T. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.