Revista de Lingüística y Lenguas Aplicadas
Vol. 13 año 2018, 99-118
EISSN 1886-6298
https://doi.org/10.4995/rlyla.2018.9075

AN EMPIRICAL STUDY ON WORD ORDER IN PREDICATES: ON SYNTAX, PROCESSING AND INFORMATION IN NATIVE AND LEARNER ENGLISH

Javier Pérez Guerra
Universidade de Vigo

Abstract: *This study focuses on the linearisation of verbal complements (or arguments) and adjuncts (or modifiers) in learner advanced English. The analysis is based on minimal pairs of sequences of complements and adjuncts in predicates, as in* He will investigate [the construction] [in a somewhat strange way which will lead to odd results] *versus* He will investigate [in a somewhat strange way which will lead to odd results] [the construction]*. Constituent linearisation has been claimed to be potentially subject to lexical, syntactic, processing and informative determinants. This paper analyses the influence of the verbal heads and the principles 'complements-first', 'end-weight' and 'given-new' on the production of predicates containing two dependents. The study has two objectives: first, to determine in which areas the ordering of such constituents in English by non-native speakers is particularly influenced by their first language (Spanish); second, to assess the plausibility of the Interface Hypothesis in a pattern which affects the so-called internal and external interfaces. Results show that internal constraints such as the length of the dependents (end-weight) and the learner's source language (i.e. compliance with complements-first) are strong predictors of ordering choices.*

Keywords: *word order, adjunct, complement, syntax, information structure, Interface Hypothesis.*

1. INTRODUCTION

This study analyses the production of syntactic structures by learners of a second language and the influence exerted by the syntax of their native language. Specifically, I investigate a number of factors which can explain the order of dependents in predicates in spoken (L2) English produced by advanced[1] learners whose native language (L1) is Spanish. The results will be contrasted with the tendencies observed in comparable spoken corpora and databases of L1 Spanish and L1 English.

Two are the objectives of this investigation. Firstly, it will be tested whether the flexibility of a given native language as regards word order (i.e. constituent order) plays a role in the design of constructions at the phrasal level in L2 productions. The L1 languages here, Spanish and English, are claimed to diverge as regards their status on a scale of word-order flexibility: Spanish, as an example of Rutherford's (1983:359) 'pragmatic-word-order' language, is considered free in this respect, whereas English, of the 'grammatical-word-order' type, is an example of a fixed word-order language (Lozano, 2014:287; Lozano and Mendikoetxea, 2008). The hypothesis will thus be that 'the status of L1 and of L2 in terms of word order is a binding condition for the learner'. In order to check the validity of such a hypothesis, I will investigate the distributional tendencies affecting the placement of adjuncts (or modifiers) and complements (arguments) in verb phrases (predicates) in spoken corpora of L1 Spanish, L1 English and non-native English produced by Spanish learners (L2 English).

Secondly, this study is inspired by the so-called Interface Hypothesis (Sorace, 2011, 2012; originally Sorace and Filiaci, 2006), which tackles the effects of 'internal' (lexical, morphosyntactic, semantic) and 'external' (pragmatic, informative) factors on the acquisition of linguistic patterns in learners, and predicts that the integration of syntactic and pragmatic conditions remains less than optimally efficient in advanced learners and gives rise to optionality. The patterns investigated here constitute an optimal testing ground for the plausibility of this claim.

[1] The learner corpus VICOLSE (see Section 3), whose only specific metadata variable which can be used to gauge the speakers' proficiency in the foreign language is 'years learning English', contains data produced by exclusively students in the third and fourth years of University degrees in English linguistics and Translation studies, which warrants its advanced level.

To cite this article: Pérez Guerra, J. (2018). "An empirical study on word order in predicates: on syntax, processing and information in native and learner English". *Revista de Lingüística y Lenguas Aplicadas*, *13*, 99-118. https://doi.org/10.4995/rlyla.2018.9075

Received: 2017-12-13 Accepted: 2018-03-07

The study is organised as follows. Section 2 outlines the case study, including the theoretical background. The methodology and corpus data are described in Section 3. The analysis of the data, including the regression model used, is the focus of Section 4. In Section 5 I discuss the results of the statistical model. Finally, Section 6 contains the main conclusions and also raises a number of issues for further research.

2. THE CASE STUDY AND THE BACKGROUND

One of the aims of this study is to investigate whether the optimal order of dependents in a phrasal construction in L1 English and in L1 Spanish predetermines its realisation by advanced Spanish speakers of L2 English. The constructions to be considered are illustrated in (1) and (2):

(1) Now I will investigate$_{V}$ [$_{COMPLEMENT}$ the construction] [$_{ADJUNCT}$ in a way which will lead to odd results].

(2) Now I will investigate [in a way which will lead to odd results] [the construction].

Whereas in (1) the complement precedes the adjunct, in (2), which in principle is the less acceptable form, the complement is sentence-final.

One of the goals of this research, as noted in Section 1, is to determine the linguistic forces that condition the placement of dependents, specifically in verb phrases in learner language. The variables, described in Section 4, will be representative of the major interfaces identifiable in the acquisition of linguistic patterns from the perspective of the Interface Hypothesis (as in Sorace, 2011). In Sorace's (2011:9) words, "there is sufficient evidence for important developmental differences between linguistic structures that require conditions of a formal nature within the grammar, and structures that require the integration of contextual factors". For merely descriptive reasons, two conditions are recognised: those pertaining to an 'internal' level, which covers processes affecting more systematic acquisition domains such as syntax, Phonological Form and Logical Form, and those assigned to an 'external' interface level, comprising performance-related tasks such as understanding (conceptual-intentional system, encompassing Christiansen and Chater's 2016 constraints from thought, cognitive and pragmatic constraints). The Interface Hypothesis predicts that the integration of internal and external conditions "remains less than optimally efficient and gives rise to optionality" (Sorace, 2011:26) with advanced learners. In Sorace's (2012:210) words, "it may be possible to determine a hierarchy of computational difficulty in highly proficient bilinguals, with structures requiring more proceduralized 'internal' mappings being less taxing than structures requiring the rapid integration and updating of contextual information". The backbone of the Interface Hypothesis is basically the recognition of two domains with different needs and consequences for language acquisition. On the one hand, acquiring the bases of the internal interface entails the adoption of the central pillars of the linguistic system of L2; on the other hand, proficiency in the so-called external interface involves mastering performance elements, and this implies a great level of linguistic delicacy in domains such as pragmatics and information structure (see, for example, Carroll *et al.,* 2000:442; Callies, 2008a:202, 2009:104). In this respect, Lozano and Callies (2018, forthcoming) claim that "[p]roperties at the external interfaces draw on information across linguistic and non-linguistic cognitive modules, so learners find it difficult to simultaneously integrate such information online as it requires more processing resources and limitations in working memory and processing capacity/efficiency are inherent features of bilingualism and SLA [Second Language Acquisition]".

The interface levels considered in the present study are: the internal one, which will be investigated through the principles of 'complements-first' (syntax) and 'end-weight' (processing), and the external level of information structure, analysed by measuring the impact of the principle 'given-new' (information) on the data. Besides, the lexical dimension will intertwine with the syntactic domain in the distinctive collexeme analysis of the data carried out in Section 4. In what follows I will describe and justify the relevance of the principles which account for the internal and external interfaces:

(i) Complements-first and the internal interface level. The syntactic determinant complements-first, so-called in, for example, Quirk *et al.* (1985:49-50) and also known as 'Arguments precede X' (Hawkins, 2007), determines that when a head holds a syntactic connection with two (or more) dependents, those acting as arguments or complements should occur next to their heads, whereas those fulfilling an adjunct or modifier role should occupy more peripheral positions within the phrases. This principle can also be taken as a consequence of Hawkins' (1999:234) 'Minimize Domains' or MiD ("[g]iven two or more categories A, B, [...] related by a grammatical rule R of combination and/or dependency, the human processor prefers to minimize the distance between them within the smallest surface structure domain sufficient for the processing of R") since, as Hawkins (2007) hypothesises, MiD is relevant especially to examples of complementation. In Second Language Acquisition, from a syntactocentric perspective (see, among others, Lardiere, 2009:176, 2012:110), accomplishment with complements-first has been treated as an issue of parameter-setting of an L2 feature, possibly different from the learner's L1 feature.

The ascription of a dependent to the class of either complement or adjunct has been controversial in the literature. As a matter of fact, a number of scholars (see, for example, Egan, 2008:2; Mair, 1990:201; Herbst *et al.,* 2004:xxxiv) have recognised that the definitions of the syntactic relations of complementation and adjunction are undesirably vague. Following Matthews (2007:187) when he says that a complement is "[a] unit in a construction either required or specifically taken by an individual member of a lexical category", this study contends that if a dependent is semantically selected or, in generative terms, subcategorised by the head category, it will be classified as a complement. By contrast, if the semantic connection between the dependent and the head is loose and, in consequence, the occurrence of the dependent in the phrase is not required by the referent of the head, then the dependent will be regarded as an adjunct. The tight semantic link holding between complements and heads is sometimes accompanied by syntactic dependencies such as lexical restrictions ('formal determination' in Greenbaum *et al.,* 1996:76, or 'syntactic integration' in Noonan, 2007:101), as in the verb *deal* plus a *with*-complement, in the adjective *keen* plus an *on*-complement, or in the noun *compliance* requiring a *with*-complement.

(ii) End-weight and the internal interface level. The second determinant potentially accounting for the placement of dependents is the principle of end-weight, as in Quirk *et al.* (1985:1398).[2] Briefly, in terms mainly of working-memory limitations, speakers and parsers (hearers) prefer orderings where the dependents which are structurally weightier (longer, heavier, structurally/syntactically more complex) tend to be placed in final positions in clauses, phrases, etc., at least in head-initial languages like English and Spanish. Precisely because Spanish is regarded as a free language in terms of word order, as Lozano and Mendikoetxea (2010:480) put it, "[l]ittle attention has been paid to the effects of [...] end-weight" under the assumption that "the relatively 'free' word order of Spanish means that weight effects may be less noticeable". On the contrary, this study will show that end-weight, far from being an unimportant determinant of linguistic ordering, is a decisive force triggering specific syntactic designs in Spanish precisely because the syntactic (internal) principle of complements-first is not as strong as in languages like English, which are more rigid in terms of word order. In fact, Lozano and Mendikoetexea (2010:480) have argued in favour of the relevance of end-weight to word order in Spanish when they consider examples like (3), and have claimed that "canonical word order with the adjunct [...] following the [heavy] objects appears to be less 'natural' than [examples] where the heavy object is in sentence-final position following the adjunct".

(3) #Vi [$_{NP}$ a los chicos de los que quería haberte contado varias historias][$_{PP}$ en el parque].
'I saw the boys I wanted to tell you stories about in the park'

If we return to (1) and (2) again, repeated here for convenience, in light of the two principles just described, (1) will be reported to be a better performance solution than (2), not only on syntactic grounds since the complement *the construction* follows the verb (and precedes the adjunct *in a somewhat strange way which will lead to odd results*), but also on processing grounds, given the amount of structure which has to be processed between the head category and the syntactic trigger[3] of the second constituent in the (local) phrase (three words in (1), i.e. *the construction in*, and twelve words in (2), i.e. *in a somewhat strange way which will lead to odd results the*).

(1) Now I will investigate [the construction] [in a somewhat strange way which will lead to odd results]

(2) Now I will investigate [in a somewhat strange way which will lead to odd results] [the construction]

(iii) Given-new and the external interface level. The notion of information packaging or information structure assumes that sentences (and phrases) are structured in order to meet the communicative/informative demands of discourse (see, among others, Vallduví and Engdahl, 1996:460). As a means of dealing with the connection between information structure and word order, this interface level will look at the informative status of arguments and adjuncts. Birner and Ward (2006:291) note that "a speaker's choice of syntactic construction can signal the constituents' assumed familiarity level", where "familiarity" (or evoked "discourse status") is understood as a measurable concept associated with the informative availability of a referent in either the preceding or the ensuing discourse.[4] Constituents

[2] End-weight has been investigated in, for example, adverb placement (Osborne, 2008), *tough*-movement (Callies, 2008b), dative alternation (Callies and Szczesniak, 2008) and verb-particle/prepositional phrase combinations (Kinne, 2017), also in (advanced) learner language.

[3] The portion of lexical structure up to the syntactic 'trigger' of the second constituent is similar to Hawkins' (2004:32) 'Constituent Recognition Domain', which is defined as follows: "[t]he [Constituent Recognition Domain] for a phrasal mother node M consists of the smallest set of terminal and non-terminal nodes that must be parsed in order to recognize M and all [Immediate Constituents] of M". This concept of syntactic trigger (a preposition in a prepositional phrase, a determiner or a possessive specifier in a noun phrase, a participle in a participial phrase, a coordinating conjunction in a coordinative structure, a subordinator or complementiser in a subordinate clause, a *wh*-proform in a relative clause, etc.) borrows Prideaux and Baker's (1986:32) concept of 'bracketing', according to which "[t]he language user assumes that when a new unit for processing is encountered or initiated, it will be marked as such". It is also heavily tied to the concept of 'incrementality', whose main assumption, in Pickering *et al.*'s (2000:5) words, is that "the language processing system must very rapidly construct a syntactic analysis for a sentence fragment, assign it a semantic interpretation". The identification of the trigger relies not only on theoretical claims (for example, the well-known notion of syntactic head in the generative literature, Frazier's 1979:43 new nodes principle, or Hawkins' 2006:209 dependency) but also on statistical information (Corley and Crocker, 2000:137).

[4] As contended by Sorace (2012:213), the relational givenness/newness of a constituent is an external property since it requires access to the previous linguistic and extralinguistic context.

are thus organised, in the unmarked cases, on a given-new continuum,[5] so that those conveying given/evoked/familiar information tend to precede those materialising informatively new discourse referents. In an attempt to assess the informative status of the dependents in the constructions under investigation here, these constituents have been classified according to the degree of textual activation of their semantic referents, that is, by looking at their preceding discourse or (local) domain for anaphoricity. The taxonomy of the dependents is as follows: (a) textually given constituents (labelled as 'g'), when the whole constituent (or at least its head noun) conveys information which is anaphoric in the preceding discourse domain (of seven or fewer sentences[6]), as in (4), from the L2 corpus VICOLSE (see Section 3), where *these two types of religion* have already been evoked in the preceding context; (b) partially given ('p') constituents, when the referent of the noun phrase is derivable (for example, via metonymy) from a previous constituent, as in (5), in which *cigarettes*, but not *smoke*, has been mentioned before; (c) deictic ('d') constituents, illustrated by *here* in (6); and (d) completely new ('n') constituents, in (7), in which the *vegetable garden* had not been referred to before.

(4) I I don't know very much *about these two types of of religion*. (VICOLSE)

(5) non-smokers complain too much about ahm (.) (tch) *about the smoke of cigarettes*. (VICOLSE)

(6) he won't come anymore *here*. (VICOLSE)

(7) I have in my house *a vegetable garden*. (VICOLSE)

After the description of the data and the methodology in Section 3, Section 4 will assess the extent to which the data comply with the principles instantiating the internal-syntax, internal-processing and external-information linguistic interfaces: respectively, complements-first, end-weight and given-new.

3. THE METHODOLOGY AND THE DATA

Research in SLA has traditionally been highly dependent on data elicitation methods and has typically disregarded data provided by learner corpora. This study advocates the use of both native and learner corpora in order to obtain information on the frequency of occurrence of the constructions under consideration. Whereas elicitation methods are welcome in the description of acquisition and can provide indispensable quantitative and qualitative information, they can offer little information on the dominant frequencies of constructional choices. Hence, as Granger (2002:7) has argued, electronic collections of authentic foreign/second language textual data, compiled "according to explicit design criteria for a particular SLA [...] purpose [,] encoded in a standardised and homogeneous way and [...] documented as to their origin of provenance", constitute vital empirical tools for frequency-driven research.[7]

Interest in so-called Learner Corpus Research is not new, and since the 70s learner corpora have been taken as bases for studies of learner interlanguage (Selinker, 1972), no doubt as a consequence of the growing availability of new learner corpora, new tools and new annotation conventions (see Callies and Paquot, 2015:162). In this respect, Granger's research and, in particular, her Contrastive Interlanguage Analysis (CIA) approach, are extremely influential. (Original) CIA, as in Granger (1996) for example, combines theoretical notions and approaches such as interlanguage (Eubank *et al.,* 1995) and so-called 'approximative linguistic systems' (Nemser, 1971) within a model in which (actual) corpus data constitute the empirical baseline.

The framework inspiring the current study, then, is the most recent version of CIA^2 (Granger, 2015:17), which reformulates some of the elements of original CIA. The model is sketched in Figure 1.

[5] The principle 'given-new' (or, 'given-before-new'), the universality of which has been defended by scholars such as Birner and Ward (2006:291), has been given a number of names and definitions in the literature (among others, Quirk *et al.*'s 1985:1357 'end-focus', Enkvist's 1987 'Old-Information-First' and Cowan's 1995:30 'old things first'). This study sticks to Quirk *et al.*'s (1985:1357) definition of the principle: "it is common to process the information in a message so as to achieve a linear presentation from low to high information value".

[6] The length of the domain (less than eight sentences) is in accordance with, for example, Hajicová and Vrbová's (1981) and Mitkov's (1998:3) proposals.

[7] Also in recent psycholinguistic research, the use of corpora to investigate learner language has been vindicated. To give an example, Monaghan and Rowland (2016:29) have claimed that "[c]orpus research has enabled us to recover the richness of the stimulus and to more effectively ascertain the available information in the environment of the language learner". In other words, corpora may provide, on the one hand, what they call "indirect evidence", not easily obtainable through elicitation tasks in a laboratory, and, on the other, frequency information which may be of paramount importance in determining learning models via statistical learning.

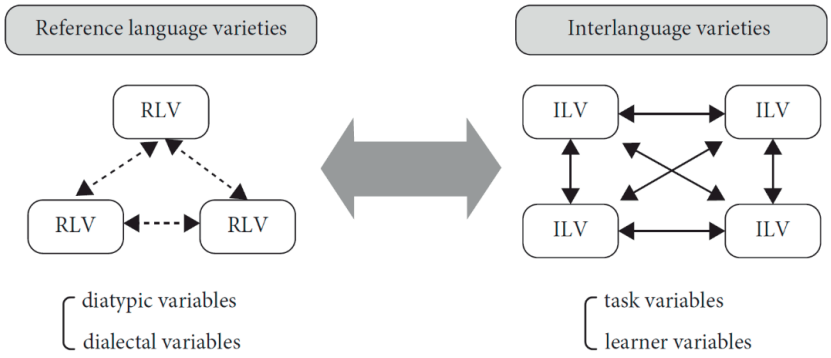

Figure 1. CIA[2].

The new version of CIA is based on bidirectional comparison/contrast between L1 and L2 varieties. The former are defined as either reference dialectal (in this study, L1 Spanish and L1 English) or diatypic varieties (permitting text-type comparability), and the latter are interlanguage varieties (instantiated here by means of learner productions).[8]

Having now summarised the theoretical and methodological background, and with the comparison/contrast between native and learner productions established as an optimal instrumental procedure to determine varietal realisation effects, I will now describe in more detail the linguistic choices and the textual sources used as empirical evidence for this study.

In order to avoid the data being influenced by the conventions of formal writing, I have opted to look exclusively at spoken data. The analysis of spoken productions, which is seen as an asset to this case study, has encountered a number of difficulties. On the one hand, as Brown *et al.* (1984:17-18) note, "the spoken language [...] consists of relatively simple sentence structures – often just sentences and incomplete sentences, strung together". On the other hand, Miller and Weinert (1998: Chapter 3), in their study of (mostly subordinate) syntactic constructions in spontaneous oral language, acknowledge that a number of these constructions, which can even be frequent in written texts, were missing from their oral data or appeared at very low percentages. As evinced by some of the frequencies provided in Section 4, two-dependent predicates are not very common in the spoken corpora used here.

Since this study focuses on L1 Spanish, L1 English and L2 English, the data have been retrieved from three spoken relevant resources, ADESSE, LOCNEC and VICOLSE, respectively. The first of these, ADESSE (*Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español*, University of Vigo, http://adesse.uvigo.es), is a syntactic database of Present-Day written and spoken L1 Spanish, comprising 1.5 million words, and has been taken as the Spanish native comparable database. Arguments (complements) and adjuncts, syntactic functions (subject, object, etc.) and syntactic categories (noun phrase, adjective phrase, clause, etc.) are annotated in ADESSE, as are other features, such as the semantic categorisation of the nouns (animate, concrete, abstract) and the arguments' semantic or thematic role (agent, theme, patient, etc). This study is based on the subcorpus of 207,948 words of spoken Spanish (produced in Spain). Second, LOCNEC (*Louvain Corpus of Native English Conversation*, Université catholique de Louvain) is a spoken collection of L1 English productions, consisting of 162,000 words, and has served here as the English native control corpus. Finally, VICOLSE (*Vigo Corpus of Learner Spoken English*, University of Vigo; Tizón-Couto, 2014) is a corpus of learner (L2) English produced by Spanish University students of English, comprising approximately 100,000 words.

A brief summary of the latter corpora seems in order here in an attempt to justify the overall comparability between the corpora of native and learner English, respectively, LOCNEC and VICOLSE. VICOLSE contains consented recordings of speech by 86 undergraduate students of English at the University of Vigo. The texts in VICOLSE are based, in terms of the medium, on oral elicitation techniques such as retelling a story, describing a picture-based real-world scene, commenting on a familiar or current topic, or giving personal opinions; with respect to the kind of tasks and variety of the data expected, the elicitation is that of a broad trawl; finally, as for genre, the learners' productions can be categorised as either narration or argumentation. The transcriptions in VICOLSE, illustrated in (8), consist of the texts and sound indications (laughing, breathing, vocalic clicks, hesitations, etc.), following the conventions adopted in LINDSEI (https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei.html).

[8] See Granger (2015) for the description of CIA[2] in response to the criticisms of original CIA, in particular, the so-called "comparative fallacy" between the levels of the model (basically L1 and L2) and the issue of the L1 norm (replaced in CIA with the concept of "reference" varieties).

(8) <stdnt 01>
[Task 1]
Little Red Riding Hood *ah* continued ahead (tch) and *ah* through the forest and then he arrived to a little house a wooden house (tch) which was her grandma's one *ahm* once she was there she knocked on the door (tch) and *ahm* her grandmother's voice said *eh* come in . she went in and *mm* she (..) kissed her grandmother and then she said *eh* here you have a gift some things I've bought for you like honey and *ah* cakes and whatever *ah* the grandmother was very happy and *ah* yeah *ah* the problem was that she was in bed cause she was quite ill and *ah* Little Red Riding Hood *ah* (..) started to ask her why was she so so ill in bed (tch) *ahm* (..) then the the grandmother said *aom* I'm very ill because *m* I have a cough (tch) and the the little *child child *started saying *mm* yeah I *??

LOCNEC collects the spoken production by 50 British (most undergraduate) Lancaster University students, aged 18-30. This corpus has been taken as the native control corpus for LINDSEI and serves here as the comparable corpus to VICOLSE. Since comparability is a major issue in CIA[2], as reported in Tizón-Couto (2014:186-190), VICOLSE and LOCNEC are analogous as far as the design of tasks and topics is concerned and practically identical as regards transcription conventions (as already pointed out, these inherited from the LINDSEI project). They obviously differ with respect to the participants since the subjects in VICOLSE are (Spanish) non-native university students of English and those in LOCNEC are British native university students, and this paves the way for linguistic comparison and contrast. Regarding modality, it must be recognised that the tasks are carried out in the form of an interview in LOCNEC, while VICOLSE contains monologistic recordings. Tizón-Couto (2014:186) undertook a careful analysis of the potential for comparability between VICOLSE and LOCNEC, here summarised in Table 1.

Table 1. Comparison of some parameters of VICOLSE and LOCNEC

Parameter	VICOLSE	LOCNEC	Degree of comparability
Speech type/elicitation	Voluntary, untimed, unprepared	Voluntary, untimed, unprepared	high
Size (words)	100,663 (76,337 words without story telling)	118,398 (B turns only)	average
Age of students	19-32	18-30	high
Year of studies	First, second, third and fourth year	Mainly first and second, also third, fourth and postgraduate	average
L2 proficiency	Intermediate–Upper intermediate–Advanced	Native	average
Tasks	Picture description, storytelling, set topic, free discussion	Picture description, set topic, free discussion	average
Topics	A favorite book or film, a famous TV show, your education, music, important decisions in your life, etc.	A country you have visited, an experience which has taught you something, a book or play you have liked/disliked	average
Genre	Monologue	Guided monologue (Dialogue)	low/average
Time of completion	2001-2008	1995-2006 (unreleased)	average

4. ANALYSIS OF THE DATA

The database for this study comprises all the examples of predicates consisting of exclusively verbs followed by two dependents (one complement and one adjunct, and vice versa) in the three electronic collections described in Section 3. In the case of ADESSE, as already noted, this was done only in the oral texts.

In order to account for the effect of the syntactic principle of complements-first in the data, the dependents were categorised into two classes: complements ('c') and adjuncts ('a'), after the application of the criteria mentioned

in Section 2. According to the order of the dependents, the examples were sorted as either 'ca' (complement plus adjunct), as in (1) above, or 'ac' (adjunct plus complement), as in (2). This dichotomous variable, with the two choices 'ca' and 'ac', has acted as the dependent (response) variable, tagged as 'var' in the statistical model.

The independent variables used in this study will now be described. First, the source of the example (corpus or database) has been annotated in the database in order to keep track of its linguistic status in terms of L1 or L2. Thus, examples retrieved from LOCNEC were encoded as L1English, those from ADESSE as L1Spanish and, finally, those from VICOLSE as L2English.

Second, the length of the dependents ('length_a' for adjuncts and 'length_c' for complements) has also been measured for the purposes of investigating the relevance of end-weight to the model. Measuring length has been approached in different ways in the literature. An influential study here is Gries (2003:83-84). In his investigation on linearisation in verb-particle predicates, Gries coded for the number of syllables as well as the number of words, and concluded that the two measures yielded very similar results, the analysis of the number of syllables being a slightly better predictor of ordering choices. Similarly, in a study on utterance length, Yaruss (1999:339) reported that "there were very strong, positive, significant correlations [...] among measures of length in words, syllables, morphemes, and clausal constituents". The same conclusion is drawn by Szmrecsányi (2004) when he compared metrics based on words, syntactic nodes in phrase-markers, and complexity counts. He concluded that "determining length in words [...] is by all means one that is nearly as accurate as the most sophisticated and cognitively, conceptually, or even psychologically 'more real' methods" (2004:1038). Similarly, Shih and Grafmiller (2011) demonstrated in their discussion of genitive and dative alternations that "the number of words [...] can act as a sufficient proxy for [...] 'weight' [length]". In the present study, the length of the subject and the objects will be evaluated through the number of words.

Third, the morphosyntactic categories of the adjuncts ('cat_a') and the complements ('cat_c') have also been coded in the database. Examples (9) to (12), from VICOLSE, illustrate complement plus adjunct constructions in which, respectively, a noun phrase (NP), a prepositional phrase (PP), an adverb phrase (AdvP) and a clause (Cl) act as complements, whilst the adjunct category is in all these cases instantiated by a prepositional phrase. By contrast, in examples (13) to (15) the adjuncts precede the complements, the former being, respectively, a noun phrase, an adverb phrase, and a prepositional phrase, and the complements being either prepositional or noun phrases.

(9) they earn [$_{NP}$ a lot of money] [$_{PP}$ with the taxes that the the the sell selling of tobacco]

(10) so he went [$_{PP}$ to the house] [$_{PP}$ with eh Little Red Riding Hood]

(11) I had to go [$_{ADVP}$ back to school] [$_{PP}$ with sixteen year-old kids]

(12) people ehm start [$_{CL}$ smoking] [$_{PP}$ in a very early age]

(13) I think that no-smokers complained [$_{NP}$ all the time] [$_{PP}$ about this this theme]

(14) he decided to wai-- wait [$_{ADVP}$ there] [$_{PP}$ to for for Little Red *Riding Hood]

(15) I have [$_{PP}$ in my house] [$_{NP}$ a vegetable garden]

Fourth, the information conveyed by the dependents is codified in the variables 'info_a' (information of the adjunct) and 'info_c' (information of the complement), in which I have opted for a dichotomous classification of the taxonomy described in Section 2 ('given', 'partially given', 'deictic' and 'new' referents) into 'given' ('g', including 'given', 'partially given' and 'deictic') and 'new' ('n').

The summary of the data is provided in Table 2. The application of a logistic regression analysis (functions 'glm' and 'lmr' in R – The R Project for Statistical Computing, https://www.r-project.org) to the previous list of variables led to the following model, in which the significant variables ($p \leq 0.01$) have been italicised in Table 3.

Table 2. Summary of the data.

Variable	Values	Data
pattern ('var')	adjunct+complement ('ac')	235
	complement+adjunct ('ca')	892
corpus	L1 English ('L1Eng')	346
	L1 Spanish ('L1 Span')	472
	L2 English ('L2Eng')	309
length of the adjunct ('length_a')	(number of words)	
length of the complement ('length_c')	(number of words)	
category of the adjunct ('cat_a')	noun phrase ('np')	78
	prepositional phrase ('pp')	581
	clause ('cl')	116
	adjective phrase ('ap')	1
	adverb phrase ('advp')	351
category of the complement ('cat_c')	noun phrase ('np')	985
	prepositional phrase ('pp')	86
	clause ('cl')	14
	adjective phrase ('ap')	21
	adverb phrase ('advp')	48
information of the adjunct ('info_a')	given ('g')	363
	new ('n')	764
information of the complement ('info_c')	given ('g')	497
	new ('n')	630

Table 3. Regression analysis.

Estimate	Std.	Error	z value	Pr(>|z|)
(Intercept)	497.051	111.122	4.473	7.71e-06
corpus[L1Span]	*-508.917*	*0.61362*	*-8.294*	*<2e-16*
corpus[L2Eng]	*-335.759*	*0.60665*	*-5.535*	*3.12e-08*
length_a	*0.42633*	*0.08723*	*4.887*	*1.02e-06*
length_c	*-0.57605*	*0.06485*	*-8.882*	*<2e-16*
cat_a[ap]	1.324.985	53.541.126	0.025	0.98026
cat_a[cl]	*135.192*	*0.44039*	*3.070*	*0.00214*
cat_a[np]	0.42676	0.40637	1.050	0.29364
cat_a[pp]	*175.962*	*0.28932*	*6.082*	*1.19e-09*
cat_c[ap]	-101.192	173.440	-0.583	0.55960
cat_c[cl]	-229.945	126.112	-1.823	0.06825
cat_c[np]	0.58460	0.95327	0.613	0.53971
cat_c[pp]	-0.92328	102.098	-0.904	0.36583
info_a[n]	-0.05156	0.24060	-0.214	0.83032
info_c[n]	*-0.65854*	*0.23983*	*-2.746*	*0.00604*

The model has proved to be adequate for the research question, instantiated by the response variable 'var' ('ca' *versus* 'ac') in a plausible way, given the results of the effect-size indexes (Nagelkerke) R^2=0.616 (understood to be very good at levels above 0.5) and C (Concordance)=0.93 (outstanding if higher than 0.9).

In order to rank the contribution of each of the variables to the overall significance of the model, Table 4 provides the values of the discrimination indexes R^2 and C of a model containing only the significant variables.

Table 4. Impact of significant variables on R^2.

Model	Variables excluded	R^2	C	Difference in R^2 (percentage)
whole model (all the variables)	(none)	0.61	0.93	(100)
only significant variables: corpus + length_a + length_c + cat_a + info_c	cat_c, info_a (not significant)	0.59	0.92	96.91
corpus + length_a + length_c + cat_a	+ info_c	0.59	0.92	96.26
corpus + length_c + cat_a + info_c	+ length_a	0.56	0.91	92.37
corpus + length_a + length_c + info_c	+ cat_a	0.55	0.91	90.58
corpus + length_a + cat_a + info_c	+ length_c	0.47	0.88	**76.78**
length_a + length_c + cat_a + info_c	+ corpus	0.43	0.86	**70.29**

The following remarks seem in order in light of the Tables 3 and 4. First, the regression analysis in Table 3 and the impact of the significant variables on the R^2 index, as set out in Table 4, show that the model is not severely affected by the drop of variables measuring either the category or the informative load of the dependents. On the one hand, the regression analysis reveals that 'cat_c' and 'info_a' are not statistically significant. On the other hand, R^2 is higher than 0.5 even if 'info_c' and 'cat_a' are removed, the effect of these variables on R^2 being, respectively, 0.65 and 5.68 percent from a model which embraces only the statistically significant variables. Second, the length of the adjunct ('length_a') does not substantially decrease the overall effect size either (only 4.54 percentage points), and this is explicable in light of the strong correlation (Spearman's rho rank correlation coefficient: 0.44, $p<0.0001$, R function 'rcorr', package Hmisc) that holds between this variable and the category of the adjunct ('cat_a') - the latter's effect size, as just pointed out, is not particularly strong. Finally, one is left with the variables accounting for the length of the complement ('length_c') and the source corpus ('corpus'), which are not only statistically very significant according to the results of the regression analysis but also have a decisive impact on the discrimination index R^2, which falls from >0.55 to 0.4 when these variables are discarded, thus reporting an effect of only 70-76 percent from the model.

Before embarking on the analysis of the two variables which have proved to be the most successful ones in explaining the overall model, namely length and source language, I will turn to the incidence of the verbs in the preference for the alternating pairs under investigation (either adjunct-complement or complement-adjunct). For this purpose, first, attention will be paid to the statistical contribution of a variable encoding the head verbs (lemmas) to the model, and, second, a distinctive collexeme analysis of the verbs favouring each of the constructional choices will be carried out. As regards the former methodological option, in an attempt to embrace or discard the random effect of the verb preceding the two dependents, a mixed-effects model (R function 'lmer', package lme4) will be applied to the data by incorporating a variable encoding the base form of the predicates' verbs. The main results of the model are shown in (16):

(16) Random effects:

Groups	Name	*Variance*	Std. Dev.
verb	(Intercept)	*0.008495*	0.09217
Residual	0.042976	0.20731	

The estimate of the variance explained by the random effect (0.0084) is very close to zero, which implies that the random effect potentially exerted by the verb preceding the dependents on the model is not statistically significant and that the regular linear model is self-reliant.

On the other hand, a so-called 'distinctive-collexeme analysis' (Gries and Stefanowitsch, 2004; Stefanowitsch and Gries, 2005) might shed some light on the strength of the preference of specific verbal heads for complement-adjunct versus adjunct-complement, or vice versa, in their predicates. In Gries and Stefanowitsch's (2004:97) words, a distinctive-collexeme analysis, couched within the family of methods of collostructional analysis, "identifies lexemes that exhibit a strong preference for one member of the pair as opposed to the other". The recognition of statistically significant bonds between a group of verbs and a specific alternating choice could be due to the form-meaning pairing of a specific syntactic option and a particular semantic class of verbs (as in, for example, Goldberg, 2006:5). This might prove the claim that learners acquire both syntax and vocabulary of the language simultaneously, as demonstrated in the psycholinguistic tradition, for example, in Scott and Fisher's (2012) case study, also on verbs. Thus, Gries' (2007) distinctive collexeme analysis has been applied only to the examples in English retrieved

from VICOLSE and LOCNEC, in an attempt to determine the existence of constructional bonds between the verbs and the response variable in native and learner English. This collostructional technique corroborated the fact that verbs do not play a major role in this model, which had already been shown by the mixed-effects results. In detail, no statistically significant correlation was observed between verbs and constructions in the native data from LOCNEC, as shown in Appendix 1, and for L2 English the results in Appendix 2 recognise significant links in only seven instances: 6 verbs (*like*, *talk*, *say*, *see*, *complain*, *know*) out of 97 prefer the adjunct-complement pattern, whereas *think* is the only verb whose correspondence with the complement-adjunct construction is statistically significant. Since these verbs account for only 59 of the examples in the database, and the significance level in those cases is not extraordinarily high (p<0.05-0.01), I have opted for discarding the analysis of the verbs.

In what follows, I will focus on the two variables that contribute most significantly to the overall model. Specifically, in order to reach conclusions on the incidence of both the length of the dependents and the corpora themselves (as the source of examples) on the preference for either the adjunct-complement or the complement-adjunct order, these two variables will be analysed in detail. First, the interaction between the variable encoding the corpus and the response variable is plotted in Figure 2, which sketches the proportions of complement-first ('ca') and complement-last ('ac') examples per corpus, that is, the degree of exemplar conformity with the internal-interface syntactic principle of complements-first.

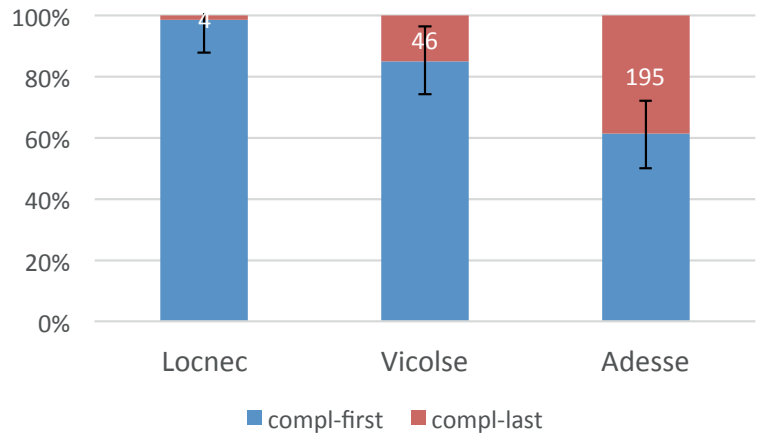

Figure 2. Corpus and construction type.

Figure 2 shows the clear decrease of the degree of compliance of the examples with complements-first on the continuum L1English-L2English-L1Spanish, such variation being statistically significant (Fisher exact test, p<0.0001). The syntactic principle of complements-first is very strong (99 percent of the examples in LOCNEC) in L1English, very strong but less so (85 percent) in L2English, and less strong (61 percent) in L1Spanish.

The second statistically significant variable was the one accounting for the length of the dependents (in particular of the complements). The boxplots in Figures 3, 4 and 5 provide the results of the length of the dependents in 'ac' and 'ca' constructions in the database in, respectively, L1 English (LOCNEC), L2 English (VICOLSE) and L1 Spanish (ADESSE).

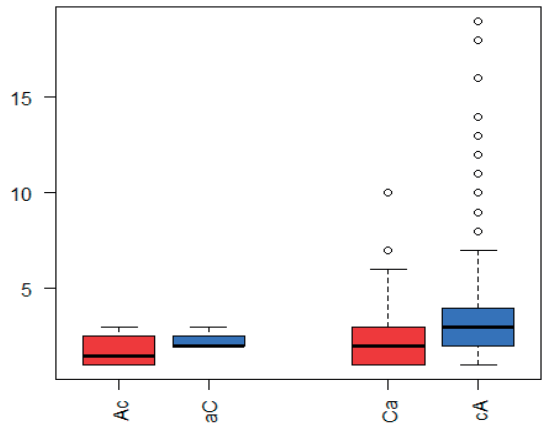

Figure 3. Length in L1 English (average number of words; first dependent in red, second dependent in blue).

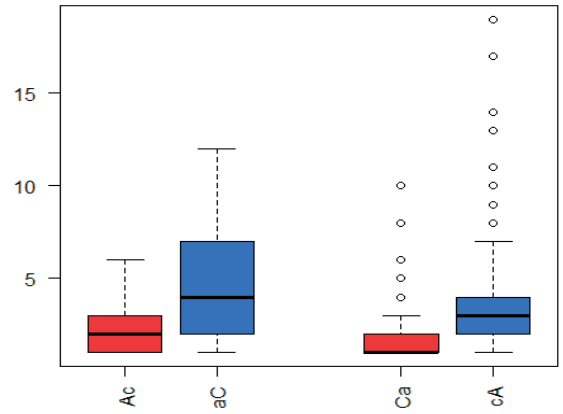

Figure 4. Length in L2 English (average number of words; first dependent in red, second dependent in blue).

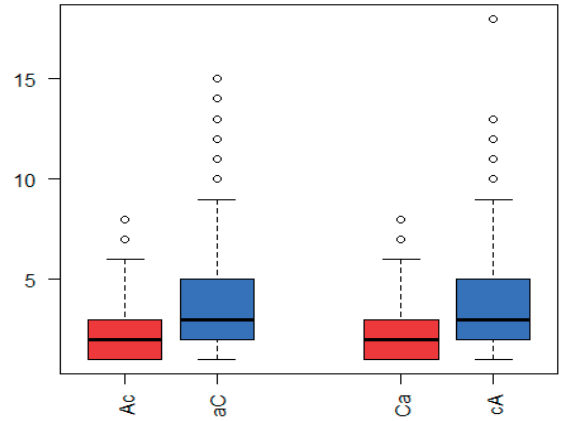

Figure 5. Length in L1 Spanish (average number of words; first dependent in red, second dependent in blue).

All these findings reveal that the length values of most of the second dependents (complements in 'ac' constructions and adjuncts in the 'ca' configuration) are greater than those of the first dependents, which corroborates the claim that end-weight is a valid principle, not only in native English and Spanish but also in the interlanguage of Spanish learners of English. That said, Figure 3 shows that in L1 English the principle of end-weight is, as already noted, reflected in the data but it is not a strong principle in the model. The core values of the length of first and second dependents, as sketched in the boxplots, do not reveal major differences. Such differences are not strong in the adjunct-complement constructions and are slightly more prominent in complement-adjunct constructions. As regards Figure 4, which portrays the values of the learner examples, end-weight is also reflected in the data and a strong determinant in the model, and is especially relevant in adjunct-complement constructions. Finally, in L1 Spanish, as plotted in Figure 5, end-weight, with which these data also comply, is a strong principle in both types of constructions, adjunct-complement and in complement-adjunct structures, at similar rates.

So, the analysis of the variable 'length' has led to the following conclusions. First, the whole database respects end-weight. Second, in L1 English end-weight is not a strong determinant and seems to operate in a subsidiary way to the syntactic principle of complements-first, since the former is seen to be respected most in adjunct-complement constructions. Thus, the data for L1 English corroborate Hawkins' (1999:232) claim that (at least) in English syntactic weight is a big predictor of relative orderings since the second dependents are longer than the first dependents in the vast majority of the two-dependent predicates investigated here. However, the empirical results give support to the view, as previously suggested by Traugott (1992:276), that "in general the light-heavy distribution [end-weight] is no longer a *major* factor in English word order" (my italics) since, according to the data, syntax, here represented by the principle of complements-first, seems to be a bigger predictor in English, in that end-weight has proved to be an ancillary determinant, one which is operative especially when complements-first is not at work, that is, in adjunct-complement predicates.

Second, in L2 English produced by speakers whose L1 is Spanish, end-weight has proved to be a stronger determining force of the distribution of dependents in predicates.[9] The data here also illustrate that end-weight is not subsidiary in L2 English to complements-first, since end-weight is in fact stronger, even in non-complement-first constructions, that is, in the adjunct-complement design.

5. DISCUSSION OF RESULTS

Section 4 provided the frequencies of complement-adjunct and adjunct-complement patterns in the three datasets (L1 Spanish, L1 English and L2 English) and accounted, on empirical grounds, for the impact of the three principles which were aimed at reflecting the interfaces, as discussed in Section 2.

The statistical analysis of the data showed that, even though complements-first and end-weight were upheld by most of the data, a cline of accomplishment with these determinants was observed depending upon the linguistic variety. The internal syntactic principle of complements-first proved to be very strong (99 percent) in L1 English, also very strong but less so (85 percent) in L2 English, and less strong (61 percent) in L1 Spanish. With respect to the internal principle of end-weight, the data revealed that this is not the major force in L1 English, is a stronger determinant in L2 English, especially in non-complement-first constructions, and constitutes a strong explanatory factor in L1 Spanish, in both complement-first and non-complement-first constructions. These results lead to the obvious conclusion that both English and Spanish internal constraints such as complements-first and end-weight exert significant influence on (L2 English) advanced learners as regards the ordering of dependents in predicates. Also and more importantly, the statistical model proved to be significantly sensitive to the application of internal factors, which led to optionally in learner language – as already pointed out in Section 4, the external constraint of given-new has not contributed to the model in a significant way. That learners are less efficient than native speakers at integrating features from the internal domains in language use is not in keeping with the Interface Hypothesis, which claims that adopting external constraints (given-new in this study) is more taxing for advanced learners than proceduralising internal principles such as complements-first and end-weight.[10]

From different theoretical standpoints, on the one hand, the Crosslinguistic Influence hypothesis (Hulk and Müller, 2000; Müller and Hulk, 2001) proposes that crosslinguistic effects are found mainly in the syntax-pragmatics interface, especially when a certain degree of superficial overlap is attested between the two languages with respect to the phenomenon. This approach is of prime importance here, since, on the one hand, word order as a strategy pertains to the syntactic domain and, on the other, most of the L1 and L2 data in English and Spanish comply with the principles of complements-first and end-weight. Accordingly, as a contribution to the Crosslinguistic Influence hypothesis, this research has shown that crosslinguistic effects can be at work in the syntactic domain when superficial overlap is attested between the two languages concerned.

On the other hand, accomplishment, on the whole, with complements-first and end-weight, and at the same time divergence between L1 and L2 as regards fulfillment of these principles is easily accommodated within theoretical proposals such as the Full Transfer/Full Access Hypothesis (Schwartz and Sprouse, 1996), which contends that the morphosyntax of L1 grammar is the initial state of L2 acquisition plus subsequent motivated restructuring because of input, learnability and Universal-Grammar considerations. In this vein, the different status of Spanish and English as regards word-order flexibility would justify not only the primacy of the syntactic principle of complements-first in L2 English over L1 Spanish but also the strength of end-weight in L2 English over L1 English during the process of interlanguage restructuring.[11]

6. CONCLUSIONS AND FURTHER RESEARCH

This paper has reported on research into word order at the phrasal level in predicates containing verbs (heads) and two dependents, one acting as complement (or argument) and another functioning as an adjunct (or modifier). The patterns have been analysed in datasets of spoken native English, spoken native Spanish and spoken learner English (produced by Spanish speakers). On theoretical grounds, this investigation has been framed within so-called Learner Corpus Research and, more specifically, Contrastive Interlanguage Analysis, which is essentially based on comparability between reference language dialectal varieties (L1 English and L1 Spanish in this case study) and interlanguage varieties (L2 English and L1 Spanish).

[9] The significant role of end-weight as an explanatory determinant of specific constructions in advanced learner language has already advocated in, for example, Callies (2006).

[10] The Interface Hypothesis has received criticism from different angles. Disentangling problematic versus unproblematic phenomena, setting boundaries between internal and external constraints, or delimiting construction-specific versus across-the-board problems, and language-specific versus universal factors have been mentioned as challenges for the Interface Hypothesis in the literature (see White, 2011 for a revision of such weaknesses).

[11] Full Transfer/Full Access is conceptually more plausible than other less absolute theoretical positions, such Rothman and Pascual y Cabo's (2014:51) when they claim that "only parts of L1 underlying syntax constitute the initial state of L2 [...], or L1 features are transferred but are underspecified, giving rise to L2 optionality in their surface reflexes".

Two interface levels have been investigated in order to determine the forces that explain the two possible linearisation choices of the dependents (adjunct-complement and complement-adjunct). Regarding the internal interface, this study has looked at compliance with the syntactic principle of complements-first, which determines that complements systematically precedes adjuncts, and at the effective fulfillment of the processing principle of end-weight, which favours designs in which longer dependents are placed in final position in their phrases. The external interface has been explored by checking the informative principle given-new (or given-before-new), according to which constituents which convey new information occur later in the clause or phrase. Since English is said to be a rigid word-order language and Spanish is characterised as a free or flexible language in terms of word order, the assumption was twofold: first, compliance with complements-first should be ensured in English and, second, end-weight and given-new should have greater presence in Spanish. From this perspective, this paper investigated the degree of fulfillment of complements-first, end-weight and given-new not only in native language but also in learner English.

The variables explored here pertained to these three principles: complements-first (internal syntactic interface), end-weight (internal processing interface) and given-new (external informative interface). The logistic regression analysis of the data and the study of the size effects of each variable contributing to the model revealed that the (given/new) information conveyed by the dependents and their morphosyntactic categorisation (noun phrases, prepositional phrases, adjective phrases, adverb phrases, clauses) have no statistically significant consequences for the preference for adjunct-complement versus complement-adjunct in the predicates. By contrast, the length of the dependents (in particular, the complements), which served to assess conformity with end-weight, and the learners' source languages were strong predictors of the response variable (adjunct-complement versus complement-adjunct). In conclusion, the data revealed that only the principles aimed at reflecting the so-called internal interface are pertinent to the study of the influence of L1 into L2 syntax. Specifically, L2 English has proved to be strongly conditioned by the core tendencies of the learners' first languages as regards compliance with the principles of complements-first and end-weight.

A number of issues remain for further research. First, the findings reported here, based on spoken linguistic productions, could be compared to similar ones using data from written databases. Second, this investigation could be enriched by looking at other two-dependent constructions, such as in noun phrases (e.g. *the author [of this book] [from the historic district of downtown Lancaster]* versus *the author [from the historic district of downtown Lancaster] [of this book]*) and adjective phrases (*keen [on music] [to a large extent]* versus *keen [to a large extent] [on music]*). Third, iconicity is another variable which might play a role in the placement of adjuncts and complements in phrases; in this respect, a fine-grained semantic analysis of the dependents might show whether semantic factors are at work or not. In this regard, well-known experiments on the order of temporal adjuncts in L1 predicates, such as Clark and Clark (1968) and Clark (1971), have determined that the meaning of the adjuncts (for example, before- versus after-adjuncts) affects their placement. Finally, the findings of the present case study could also be complemented with on-line tasks testing the subjects' acceptability judgements of the corpus examples used here.

ACKNOWLEDGEMENTS

I am grateful to the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (grant no. FFI2016-77018-P) and Xunta de Galicia (grant no. ED431C 2017/50) for generous financial support.

REFERENCES

Birner, B., Ward, G. L. (2006). "Information structure". In B. Aarts and A. McMahon (Eds.), *The handbook of English linguistics*, *1*, 291–317. Oxford: Blackwell. https://doi.org/10.1002/9780470753002.ch13

Brown, G., Anderson, A., Shillcock, R., Yule, G. (1984). *Teaching talk: strategies for production and assessment*. Cambridge: Cambridge University Press.

Callies, M. (2006). "Why money can't buy you anything in German: a functional-typological approach to the mapping of semantic roles to syntactic functions in SLA". In J. Arabski (Ed.), *Cross-linguistic influences in the second language lexicon*, 111–129. Clevedon: Multilingual Matters.

Callies, M. (2008a). "Easy to understand but difficult to use? Raising constructions and information packaging in the advanced learner variety". In G. Gilquin, S. Papp and M. B. Díez-Bedmar (Eds.), *Linking contrastive and learner corpus research*, *66*, 201-226. Amsterdam: Rodopi. https://doi.org/10.1163/9789401206204_009

Callies, M. (2008b). "Argument realization and information packaging in *tough*-movement constructions – a learner-corpus-based investigation". In D. Gabryś-Barker (Ed.), *Morphosyntactic issues in Second Language Acquisition studies*, 29-46. Clevedon: Multilingual Matters.

Callies, M. (2009). *Information highlighting in advanced learner English. The syntax-pragmatics interface in Second Language Acquisition*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.186

Callies, M., Paquot, M. (2015). “An interview with Yukio Tono”. *International Journal of Learner Corpus Research*, *1*/1, 160-171. https://doi.org/10.1075/ijlcr.1.1.06lee

Callies, M., Szczesniak, K. (2008). "Argument realization, information status and syntactic weight – a learner-corpus study of the dative alternation". In P. Grommes and M. Walter (Eds.), *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*, 165-187. Tübingen: Niemeyer.

Carroll, M., Murcia-Serra, J. *et al.* (2000). “The relevance of information organization to Second Language Acquisition studies. The descriptive discourse of advanced adult learners of German”. *Studies in Second Language Acquisition*, *22*/3, 441-466. https://doi.org/10.1017/S0272263100003065

Christiansen, M. H., Chater, N. (2016). *Creating language. Integrating evolution, acquisition, and processing*. Cambridge, Mass.: MIT Press.

Clark, E. V. (1971). “On the acquisition of the meaning of before and after”. *Journal of Verbal Learning and Verbal Behavior*, *10*/3, 266-275. https://doi.org/10.1016/S0022-5371(71)80054-3

Clark, H. H., Clark, E. V. (1968). “Semantic distinctions and memory for complex sentences”. *Quarterly Journal of Experimental Psychology*, *20*/2, 129-138. https://doi.org/10.1080/14640746808400141

Corley, S., Crocker, M. W. (2000). "The modular statistical hypothesis: exploring lexical category ambiguity". In M. W. Crocker, M. Pickering and Ch. Clifton Jr. (Eds.), *Architectures and mechanisms for language processing*, 135–160. Cambridge: Cambridge University Press.

Cowan, R. (1995)."What are discourse principles made of?" In P. A. Downing and M. Noonan (Eds.), *Word order in discourse*, 29-49. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.30.03cow

Egan, Th. (2008). *Non-finite complementation. A usage-based study of infinitive and* -ing *clauses in English*. Amsterdam: Rodopi. https://doi.org/10.1163/9789401205542

Enkvist, N. E. (1987). "Text linguistics for the applier: an orientation". In U. Connor and R. B. Kaplan (Eds.), *Writing across languages: analysis of L2 texts*, 23-42. Reading, MA: Addison-Wesley.

Eubank, L., Selinker, L., Sharwood Smith, M. (1995). *The current state of interlanguage*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.73

Frazier, L. (1979). *On comprehending sentences: syntactic parsing strategies*. Phd thesis. University of Connecticut.

Goldberg, A. E. (2006). *Constructions at work: the nature of generalizations in language*. Oxford: Oxford University Press.

Granger, S. (1996). "From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora". In K. Aijmer, B. Altenberg and M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies*, 37-51. Lund: Lund University Press.

Granger, S. (2002). "A bird’s eye of learner corpus research". In S. Granger, J. Hung and S. Petch-Tyson (Eds.), *Computer learner corpora, Second Language Acquisition and Foreign Language Teaching*, 3-33. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.6.04gra

Granger, S. (2015). “Contrastive Interlanguage Analysis. A reappraisal”. *International Journal of Learner Corpus Research*, *1*/1, 7–24. https://doi.org/10.1075/ijlcr.1.1.01gra

Greenbaum, S., Nelson, G., Weitzman, M. (1996). "Complement clauses in English". In J. Thomas and M. H. Short (Eds.), *Using corpora for language research: studies in honour of Geoffrey Leech*, 76-91. London: Longman.

Gries, S. Th. (2003). *Multifactorial analysis in corpus linguistics: a study of particle placement*. New York: Continuum.

Gries, S. Th. (2007). Coll.analysis 3.2a. A program for R for Windows 2.x [Computer software].

Gries, S. Th., Stefanowitsch, A. (2004). “Extending collostructional analysis. A corpus-based perspective on ‘alternations’”. *International Journal of Corpus Linguistics*, *9*/1, 97-129. https://doi.org/10.1075/ijcl.9.1.06gri

Hajicová, E., Vrbová, J. (1981). “On the saliency of the elements of the stock of shared knowledge”. *Folia Linguistica*, *15*, 291–303. https://doi.org/10.1515/flin.1981.15.3-4.291

Hawkins, J. A. (1999). “The relative order of prepositional phrases in English: going beyond manner-place-time”. *Language Variation and Change*, *11*/3, 231-266. https://doi.org/10.1017/S0954394599113012

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hawkins, J. A. (2006). "Gradedness as relative efficiency in the processing of syntax and semantics". In G. Fanselow, C. Féry, R. Vogel and M. Schlesewsky (Eds.), *Gradience in grammar. Generative perspectives*, 207-226. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199274796.003.0011

Hawkins, J. A. (2007). Performance and grammatical variation in the ordering of verb, direct objects and obliques. Paper presented at DGfS, Siegen.

Herbst, Th., Heath, D., Roe, I. F., Götz, D., Klotz, M. (2004). *A valency dictionary of English. A corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110892581

Hulk, A., Müller, N. (2000). "Bilingual first language acquisition at the interface between syntax and pragmatics". *Bilingualism: Language and Cognition*, *3*/3, 227-244. https://doi.org/10.1017/S1366728900000353

Kinne, A. (2017). "Weight as a determinant of syntactic variation in English L1 and L2 academic writing: a corpus study of weight effects in verb-particle/prepositional phrase combinations". In P. De Haan, S. van Vuuren and R. de Vries (Eds.), *Language, learners and levels: progression and variation. Corpora and language in use – Proceedings, 3*, 129–151. Louvain-la-Neuve: Presses universitaires de Louvain.

Lardiere, D. (2009). "Some thoughts on the contrastive analysis of features in Second Language Acquisition". *Second Language Research*, *25*/2, 173-227. https://doi.org/10.1177/0267658308100283

Lardiere, D. (2012). "Linguistic approaches to second language syntax". In S. M. Gass and A. Mackey (Eds.), *The Routledge handbook of Second Language Acquisition*, 106-126. London: Routledge.

Lozano, C. (2014). "Word order in second language Spanish". In K. L. Geeslin (Ed.), *The handbook of Spanish Second Language Acquisition*, 287-310. Malden, MA: Wiley-Blackwell.

Lozano, C., Callies, M. (2018, forthcoming). "Word order and information structure in advanced SLA". In P. Malovrh and A. G. Benati (Eds.), *The handbook of advanced proficiency in Second Language Acquisition*. Oxford: Wiley-Blackwell.

Lozano, C., Mendikoetxea, A. (2008). "Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: a corpus analysis". In G. Gilquin, S. Papp and M. B. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research*, 85-125. Amsterdam: Rodopi. https://doi.org/10.1163/9789401206204_005

Lozano, C., Mendikoetxea, A. (2010). "Interface conditions on postverbal subjects: a corpus study of L2 English". *Bilingualism: Language and Cognition*, *13*/4, 475-497. https://doi.org/10.1017/S1366728909990538

Mair, Ch. (1990). *Infinitival complement clauses in English*. Cambridge: Cambridge University Press.

Matthews, P. H. (2007). *Syntactic relations. A critical survey*. Cambridge: Cambridge University Press.

Miller, J. E., Weinert, R. (1998). *Spontaneous spoken language: syntax and discourse*. London: Clarendon Press.

Mitkov, R. (1998). "The latest in anaphora resolution: going multilingual". *SEPLN*, *23*, 1-17.

Monaghan, P., Rowland, C. F. (2016). "Combining language corpora with experimental and computational approaches for language acquisition research". *Language Learning*, *67*/1, 14-39. https://doi.org/10.1111/lang.12221

Müller, N., Hulk, A. (2001). "Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages". *Bilingualism: Language and Cognition*, *4*/1, 1–21. https://doi.org/10.1017/S1366728901000116

Nemser, W. (1971). "Approximative systems of foreign language learners". *International Review of Applied Linguistics*, *9*/2, 115-123. https://doi.org/10.1515/iral.1971.9.2.115

Noonan, M. (2007 [1985]). "Complementation". In T. Shopen (Ed.), *Language typology and syntactic description. Volume II: Complex constructions*, 52-150. Cambridge: Cambridge University Press.

Osborne, J. (2008). "Adverb placement in post-intermediate learner English: a contrastive study of learner corpora". In G. Gilquin, M. B. Diez-Bedmar and S. Papp (Eds.), *Linking contrastive and Learner Corpus Research*, 127–146. Amsterdam: Rodopi. https://doi.org/10.1163/9789401206204_006

Pickering, M. J., Clifton Jr., Ch., Crocker, M. W. (2000). "Architectures and mechanisms in sentence comprehension". In M. W. Crocker, M. Pickering and Ch. Clifton Jr. (Eds.), *Architectures and mechanisms for language processing*, 1–28. Cambridge: Cambridge University Press.

Prideaux, G. D., Baker, W. J. (1986). *Strategies and structures: the processing of relative clauses*. Amsterdam: John Benjamins.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Rothman, J., Pascual y Cabo, D. (2014). "Generative approaches to Spanish Second Language Acquisition". In K. L. Geeslin (Ed.), *The Handbook of Spanish Second Language Acquisition*, 46-63. Malden, MA: Wiley-Blackwell.

Rutherford, W. E. (1983). "Language typology and language transfer". In S. M. Gass and L. Selinker (Eds.), *Language transfer in language learning*, 358-370. Rowley, Mass.: Newbury House.

Schwartz, B. D., Sprouse, R. A. (1996). "L2 cognitive states and the Full Transfer/Full Access model". *Second Language Research*, *12*, 40-72. https://doi.org/10.1177/026765839601200103

Scott, R. M., Fisher, C. (2012). "2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty". *Cognition*, *122*/2, 163-180. https://doi.org/10.1016/j.cognition.2011.10.010

Selinker, L. (1972). "Interlanguage". *International Review of Applied Linguistics in Language Teaching*, *10*/1–4, 209-232. https://doi.org/10.1515/iral.1972.10.1-4.209

Shih, S., Grafmiller, J. (2011). Weighing in on end weight. Paper presented at the 85th Annual Meeting of the Linguistic Society of America. Pittsburg, PA, 9–11 January.

Sorace, A. (2011). "Pinning down the concept of 'interface' in bilingualism". *Linguistic Approaches to Bilingualism*, *1*/1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A. (2012). "Pinning down the concept of 'interface' in bilingualism: a reply to peer commentaries". *Linguistic Approaches to Bilingualism*, *2*/2, 209-216. https://doi.org/10.1075/lab.2.2.04sor

Sorace, A., Filiaci, F. (2006). "Anaphora resolution in near-native speakers of Italian". *Second Language Research*, *22*/3, 339-368. https://doi.org/10.1191/0267658306sr271oa

Stefanowitsch, A., Gries, S. Th. (2005). "Covarying collexemes". *Corpus Linguistics and Linguistic Theory*, *1*/1, 1-43. https://doi.org/10.1515/cllt.2005.1.1.1

Szmrecsányi, B. M. (2004). "On operationalizing syntactic complexity". In G. Purnelle, C. Fairon and A. Dister (Eds.), *Le poids des mots: Proceedings of the 7th International Conference on Textual Data Statistical Analysis, March 10-12, 2004, Vol II,* 1032–1039. Louvain-la-Neue: Presses universitaires de Louvain.

Tizón-Couto, B. (2014). *Clausal complements in native and learner spoken English. A corpus-based study with Lindsei and Vicolse*. Bern: Peter Lang. https://doi.org/10.3726/978-3-0351-0428-8

Traugott, E. C. (1992). "Syntax". In R. M. Hogg (Ed.), *The Cambridge history of the English language. Volume I: The beginnings to 1066*, 168-289. Cambridge: Cambridge University Press.

Vallduví, E., Engdahl, E. (1996). "The linguistic realization of information packaging". *Linguistics*, *34*, 459-519. https://doi.org/10.1515/ling.1996.34.3.459

White, L. (2011). "Second Language Acquisition at the interfaces". *Lingua*, *121*/4: 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

Yaruss, J. S. (1999). "Utterance length, syntactic complexity, and childhood stuttering". *Journal of Speech, Language, and Hearing Research*, *42*/2, 329–344. https://doi.org/10.1044/jslhr.4202.329

APPENDIX 1.

Distinctive collexeme analysis of the LOCNEC dataset (*obs(ac/ca)*: observed frequency of adjunct-complement/complement-adjunct; *exp(ac/ca)*: expected frequency; *SumAbsDev*: sum of the absolute value of the log-transformed *p*-values of the words/constructions to be contrasted (the larger, the stronger the deviation); *fav_cxn*: favoured construction by each verb, as shown by its deviation from the expected frequency; *abs(pbin)*: absolute value of the log-transformed *p*-values of the words/constructions to be contrasted; *significance*: $p<0.001$ if abs(pbin)>3, $p<0.01$ if abs(pbin)>2, $p<0.05$ if abs(pbin)>1.30103, *no* if abs(pbin)≤1.30103)

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
write	1	4	0.0578	4.9422	2.4962	ac	1.2481	no
stay	1	5	0.0694	5.9306	2.3428	ac	1.1714	no
be	2	39	0.474	40.526	2.1776	ac	1.0888	no
do	0	41	0.474	40.526	0.414	ca	0.207	no
go	0	36	0.4162	35.5838	0.3636	ca	0.1818	no
have	0	32	0.3699	31.6301	0.3232	ca	0.1616	no
see	0	17	0.1965	16.8035	0.1716	ca	0.0858	no
show	0	12	0.1387	11.8613	0.1212	ca	0.0606	no
live	0	11	0.1272	10.8728	0.111	ca	0.0555	no
get	0	10	0.1156	9.8844	0.101	ca	0.0505	no
spend	0	9	0.104	8.896	0.0908	ca	0.0454	no
take	0	8	0.0925	7.9075	0.0808	ca	0.0404	no
teach	0	6	0.0694	5.9306	0.0606	ca	0.0303	no
dance	0	4	0.0462	3.9538	0.0404	ca	0.0202	no
enjoy	0	4	0.0462	3.9538	0.0404	ca	0.0202	no
put	0	4	0.0462	3.9538	0.0404	ca	0.0202	no
choose	0	3	0.0347	2.9653	0.0302	ca	0.0151	no

Appendix 1, continues in next page

Appendix 1, continues from the previous page

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
forget	0	3	0.0347	2.9653	0.0302	ca	0.0151	no
like	0	3	0.0347	2.9653	0.0302	ca	0.0151	no
move	0	3	0.0347	2.9653	0.0302	ca	0.0151	no
advertise	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
apply	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
come	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
drive	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
feel	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
find	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
keep	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
learn	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
need	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
paint	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
read	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
set	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
speak	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
study	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
visit	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
watch	0	2	0.0231	1.9769	0.0202	ca	0.0101	no
apply	0	1	0.0116	0.9884	0.01	ca	0.005	no
arrange	0	1	0.0116	0.9884	0.01	ca	0.005	no
ask	0	1	0.0116	0.9884	0.01	ca	0.005	no
beat	0	1	0.0116	0.9884	0.01	ca	0.005	no
blame	0	1	0.0116	0.9884	0.01	ca	0.005	no
build	0	1	0.0116	0.9884	0.01	ca	0.005	no
buy	0	1	0.0116	0.9884	0.01	ca	0.005	no
catch	0	1	0.0116	0.9884	0.01	ca	0.005	no
charge	0	1	0.0116	0.9884	0.01	ca	0.005	no
clean	0	1	0.0116	0.9884	0.01	ca	0.005	no
consider	0	1	0.0116	0.9884	0.01	ca	0.005	no
curl	0	1	0.0116	0.9884	0.01	ca	0.005	no
cut	0	1	0.0116	0.9884	0.01	ca	0.005	no
describe	0	1	0.0116	0.9884	0.01	ca	0.005	no
distance	0	1	0.0116	0.9884	0.01	ca	0.005	no
earn	0	1	0.0116	0.9884	0.01	ca	0.005	no
ease	0	1	0.0116	0.9884	0.01	ca	0.005	no
eat	0	1	0.0116	0.9884	0.01	ca	0.005	no
envisage	0	1	0.0116	0.9884	0.01	ca	0.005	no
fill	0	1	0.0116	0.9884	0.01	ca	0.005	no
finish	0	1	0.0116	0.9884	0.01	ca	0.005	no
flatter	0	1	0.0116	0.9884	0.01	ca	0.005	no
fulfil	0	1	0.0116	0.9884	0.01	ca	0.005	no
hand	0	1	0.0116	0.9884	0.01	ca	0.005	no
has	0	1	0.0116	0.9884	0.01	ca	0.005	no
hear	0	1	0.0116	0.9884	0.01	ca	0.005	no
hold	0	1	0.0116	0.9884	0.01	ca	0.005	no

Appendix 1, continues in next page

Appendix 1, continues from the previous page

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
immerse	0	1	0.0116	0.9884	0.01	ca	0.005	no
introduce	0	1	0.0116	0.9884	0.01	ca	0.005	no
judge	0	1	0.0116	0.9884	0.01	ca	0.005	no
leave	0	1	0.0116	0.9884	0.01	ca	0.005	no
lose	0	1	0.0116	0.9884	0.01	ca	0.005	no
love	0	1	0.0116	0.9884	0.01	ca	0.005	no
make	0	1	0.0116	0.9884	0.01	ca	0.005	no
meet	0	1	0.0116	0.9884	0.01	ca	0.005	no
miss	0	1	0.0116	0.9884	0.01	ca	0.005	no
narrow	0	1	0.0116	0.9884	0.01	ca	0.005	no
nudge	0	1	0.0116	0.9884	0.01	ca	0.005	no
phone	0	1	0.0116	0.9884	0.01	ca	0.005	no
play	0	1	0.0116	0.9884	0.01	ca	0.005	no
prop	0	1	0.0116	0.9884	0.01	ca	0.005	no
rent	0	1	0.0116	0.9884	0.01	ca	0.005	no
say	0	1	0.0116	0.9884	0.01	ca	0.005	no
seem	0	1	0.0116	0.9884	0.01	ca	0.005	no
sell	0	1	0.0116	0.9884	0.01	ca	0.005	no
send	0	1	0.0116	0.9884	0.01	ca	0.005	no
show	0	1	0.0116	0.9884	0.01	ca	0.005	no
spoil	0	1	0.0116	0.9884	0.01	ca	0.005	no
start	0	1	0.0116	0.9884	0.01	ca	0.005	no
store	0	1	0.0116	0.9884	0.01	ca	0.005	no
tell	0	1	0.0116	0.9884	0.01	ca	0.005	no
think	0	1	0.0116	0.9884	0.01	ca	0.005	no
travel	0	1	0.0116	0.9884	0.01	ca	0.005	no
use	0	1	0.0116	0.9884	0.01	ca	0.005	no
walk	0	1	0.0116	0.9884	0.01	ca	0.005	no
waste	0	1	0.0116	0.9884	0.01	ca	0.005	no

APPENDIX 2. DISTINCTIVE COLLEXEME ANALYSIS OF THE VICOLSE DATASET

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
like	4	1	0.7443	4.2557	5.3298	ac	2.6649	$p<0.01$
talk	4	1	0.7443	4.2557	5.3298	ac	2.6649	$p<0.01$
say	5	4	1.3398	7.6602	4.5284	ac	2.2642	$p<0.01$
see	0	28	4.1683	23.8317	3.9202	ca	1.9601	$p<0.05$
think	3	1	0.5955	3.4045	3.862	ac	1.931	$p<0.05$
complain	2	0	0.2977	1.7023	3.3088	ac	1.6544	$p<0.05$
know	3	3	0.8932	5.1068	2.6668	ac	1.3334	$p<0.05$
look	2	1	0.4466	2.5534	2.4454	ac	1.2227	no
eat	0	17	2.5307	14.4693	2.38	ca	1.19	no
go	4	7	1.6375	9.3625	2.337	ac	1.1685	no
begin	1	0	0.1489	0.8511	1.6544	ac	0.8272	no
bring	1	0	0.1489	0.8511	1.6544	ac	0.8272	no

Appendix 2, continues in next page

Appendix 2, continues from the previous page

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
meet	1	0	0.1489	0.8511	1.6544	ac	0.8272	no
miss	1	0	0.1489	0.8511	1.6544	ac	0.8272	no
wait	1	0	0.1489	0.8511	1.6544	ac	0.8272	no
leave	2	3	0.7443	4.2557	1.5772	ac	0.7886	no
hear	0	11	1.6375	9.3625	1.54	ca	0.77	no
help	0	8	1.1909	6.8091	1.12	ca	0.56	no
come	1	1	0.2977	1.7023	1.1196	ac	0.5598	no
enjoy	1	1	0.2977	1.7023	1.1196	ac	0.5598	no
ask	0	7	1.0421	5.9579	0.98	ca	0.49	no
do	0	6	0.8932	5.1068	0.84	ca	0.42	no
listen	0	6	0.8932	5.1068	0.84	ca	0.42	no
learn	1	2	0.4466	2.5534	0.8326	ac	0.4163	no
give	0	5	0.7443	4.2557	0.7	ca	0.35	no
study	1	11	1.7864	10.2136	0.6976	ca	0.3488	no
spend	1	3	0.5955	3.4045	0.6462	ac	0.3231	no
follow	0	4	0.5955	3.4045	0.56	ca	0.28	no
remember	0	4	0.5955	3.4045	0.56	ca	0.28	no
have	4	26	4.466	25.534	0.549	ca	0.2745	no
start	1	4	0.7443	4.2557	0.514	ac	0.257	no
take	1	4	0.7443	4.2557	0.514	ac	0.257	no
find	0	3	0.4466	2.5534	0.42	ca	0.21	no
finish	0	3	0.4466	2.5534	0.42	ca	0.21	no
send	0	3	0.4466	2.5534	0.42	ca	0.21	no
smell	0	3	0.4466	2.5534	0.42	ca	0.21	no
tell	0	3	0.4466	2.5534	0.42	ca	0.21	no
touch	0	3	0.4466	2.5534	0.42	ca	0.21	no
put	1	7	1.1909	6.8091	0.3598	ca	0.1799	no
believe	0	2	0.2977	1.7023	0.28	ca	0.14	no
carry	0	2	0.2977	1.7023	0.28	ca	0.14	no
change	0	2	0.2977	1.7023	0.28	ca	0.14	no
explain	0	2	0.2977	1.7023	0.28	ca	0.14	no
hide	0	2	0.2977	1.7023	0.28	ca	0.14	no
introduce	0	2	0.2977	1.7023	0.28	ca	0.14	no
invite	0	2	0.2977	1.7023	0.28	ca	0.14	no
make	0	2	0.2977	1.7023	0.28	ca	0.14	no
push	0	2	0.2977	1.7023	0.28	ca	0.14	no
set	0	2	0.2977	1.7023	0.28	ca	0.14	no
use	0	2	0.2977	1.7023	0.28	ca	0.14	no
become	0	1	0.1489	0.8511	0.14	ca	0.07	no
cause	0	1	0.1489	0.8511	0.14	ca	0.07	no
chase	0	1	0.1489	0.8511	0.14	ca	0.07	no
choose	0	1	0.1489	0.8511	0.14	ca	0.07	no
compare	0	1	0.1489	0.8511	0.14	ca	0.07	no
condition	0	1	0.1489	0.8511	0.14	ca	0.07	no
confront	0	1	0.1489	0.8511	0.14	ca	0.07	no
continue	0	1	0.1489	0.8511	0.14	ca	0.07	no

Appendix 2, continues in next page

Appendix 2, continues from the previous page

verb	obs(ac)	obs(ca)	exp(ac)	exp(ca)	SumAbsDev	fav_cxn	abs(pbin)	significance
cook	0	1	0.1489	0.8511	0.14	ca	0.07	no
create	0	1	0.1489	0.8511	0.14	ca	0.07	no
cut	0	1	0.1489	0.8511	0.14	ca	0.07	no
develop	0	1	0.1489	0.8511	0.14	ca	0.07	no
download	0	1	0.1489	0.8511	0.14	ca	0.07	no
earn	0	1	0.1489	0.8511	0.14	ca	0.07	no
educate	0	1	0.1489	0.8511	0.14	ca	0.07	no
employ	0	1	0.1489	0.8511	0.14	ca	0.07	no
fail	0	1	0.1489	0.8511	0.14	ca	0.07	no
get	0	1	0.1489	0.8511	0.14	ca	0.07	no
ignore	0	1	0.1489	0.8511	0.14	ca	0.07	no
keep	0	1	0.1489	0.8511	0.14	ca	0.07	no
kill	0	1	0.1489	0.8511	0.14	ca	0.07	no
kiss	0	1	0.1489	0.8511	0.14	ca	0.07	no
live	0	1	0.1489	0.8511	0.14	ca	0.07	no
love	0	1	0.1489	0.8511	0.14	ca	0.07	no
mislead	0	1	0.1489	0.8511	0.14	ca	0.07	no
mix	0	1	0.1489	0.8511	0.14	ca	0.07	no
notice	0	1	0.1489	0.8511	0.14	ca	0.07	no
offer	0	1	0.1489	0.8511	0.14	ca	0.07	no
open	0	1	0.1489	0.8511	0.14	ca	0.07	no
pick	0	1	0.1489	0.8511	0.14	ca	0.07	no
practice	0	1	0.1489	0.8511	0.14	ca	0.07	no
prefer	0	1	0.1489	0.8511	0.14	ca	0.07	no
publish	0	1	0.1489	0.8511	0.14	ca	0.07	no
quit	0	1	0.1489	0.8511	0.14	ca	0.07	no
rescue	0	1	0.1489	0.8511	0.14	ca	0.07	no
sell	0	1	0.1489	0.8511	0.14	ca	0.07	no
shoot	0	1	0.1489	0.8511	0.14	ca	0.07	no
sit	0	1	0.1489	0.8511	0.14	ca	0.07	no
stand	0	1	0.1489	0.8511	0.14	ca	0.07	no
stay	0	1	0.1489	0.8511	0.14	ca	0.07	no
stop	0	1	0.1489	0.8511	0.14	ca	0.07	no
taste	0	1	0.1489	0.8511	0.14	ca	0.07	no
teach	0	1	0.1489	0.8511	0.14	ca	0.07	no
throw	0	1	0.1489	0.8511	0.14	ca	0.07	no
treat	0	1	0.1489	0.8511	0.14	ca	0.07	no
understand	0	1	0.1489	0.8511	0.14	ca	0.07	no
value	0	1	0.1489	0.8511	0.14	ca	0.07	no