

Wang, Yang. 2016. *The Idiom Principle and L1 Influence. A Contrastive Learner-corpus Study of Delexical Verb + Noun Collocations*. Amsterdam and Philadelphia: John Benjamins Publishing Company. Studies in Corpus Linguistics 77. 249 pages.

The monograph *The Idiom Principle and L1 Influence. A contrastive learner-corpus study of delexical verb + noun collocations* is a publication in the series Studies in Corpus Linguistics. As the title indicates, the study focuses on the frequencies and patterns of delexical verbs + nouns collocations, and the aim is to investigate whether Chinese and Swedish learners of English use such constructions in the same way that native speakers (NS) do. The study also includes L1 corpora in Chinese and Swedish, which accounts for the reference to L1 influence in the title: the author aims to identify the role of L1 transfer in the potential underuse or overuse of the learners.

The book is organized into seven chapters, plus references and four appendices. Chapter 1, entitled 'Introduction', clearly and concisely describes the background for the study, situating it within the tradition of corpus-based studies of interlanguage, with obvious links to the field of second language acquisition (SLA). A central issue in the investigation is the extent to which L2 learners rely on prefabricated chunks or individual words as "the 'building blocks' of language" (p. 4), and the author gives an overview of different views on the processing mechanisms and procedures employed by language learners. The author states that delexical verb + noun collocations are particularly useful when it comes to shedding light on the approaches learners may take: do they use the "open-choice principle" and treat such constructions as a series of open slots which can be filled according to the writer's choice, or do they follow the "idiom principle", and treat such constructions as prefabricated units? Given that many combinations are idiomatic, while some allow for choice, the answer will probably prove to be that learners do both, but, as the author concludes, "[t]he liberty that the learners take with the use of such word combinations will therefore shed light on how the two principles work together in learner language" (p. 6).

The author goes on to discuss the fact that delexical collocations exist in many languages, and that this means they are susceptible to L1 influence. However, given the relative difference in typological closeness between English and Swedish, and between English and Chinese, it is

likely that transfer may influence the two learner groups under scrutiny rather differently.

Next, there is a brief description of the Integrated Contrastive Model, which is the analytical framework employed in the design of the study.

The final section of chapter 1, not counting the overview of the structure of the book, is devoted to an overview of English teaching in China and Sweden, on the basis that such factors as “the status of English, teaching methodology, and the amount and nature of TL input in the two countries” may be “important in accounting for the learner performance in the study” (p. 10). The upshot is that although English is a foreign language in both countries, Swedish learners get far more exposure to authentic English input, and this will allow for a study of how environmental factors affect learner output.

Chapter 2 presents the data and methodology employed in the investigation. The material comprises data from two learner corpora, the Uppsala Student English Corpus (USE) and the Corpus of Chinese Learner English (CCLE), from which argumentative essays were selected (253 from USE and 230 from CCLE). To serve as NS control corpus, 343 newspaper editorials from FLOB and FROWN were chosen. In addition, two smaller corpora of L1 Swedish (approximately 38,000 words) and L1 Chinese (approximately 45,000 words) were used. The author recognizes the limitation involved in using such small corpora, stating that “[t]he limited amount of L1 material means that it is sometimes difficult to draw statistically valid conclusions involving the L1s”.

From these corpora, the delexical verb + noun collocations were extracted using the *WordSmith Concordancer*. Six high-frequency verbs were included: ‘have’, ‘get’, ‘make’, ‘give’, and ‘take’. The author gives a thorough account of the criteria used to identify the collocations included in the study, and there are a number of useful examples. There is also a description of the L1 Swedish and Chinese counterparts of the English verbs.

In chapter 3 the results relating to frequency of delexical verb + noun collocations are presented. The chapter starts with an overview of the role of frequency and frequency information in language learning, and then goes on to describe the quantitative results. There is a very brief note on the statistical testing that has been performed, and the results are then visualized through (primarily) bar charts and tables showing the

frequencies for each corpus, first for the collocations overall, and then for the individual verb + noun collocations. The discussion is illustrated with examples, and the main findings are that both learner varieties overuse delexical verb + noun collocations overall, but they have different preferences regarding the high-frequency verbs. The overuse seems to be primarily due to developmental factors (p. 52), and there is some evidence that the “learners’ perceptions about language distance can either trigger or constrain transfer” (p. 53).

Chapter 4 comprises a qualitative analysis, intended to complement the quantitative analysis presented in chapter 3: the quantitative results showed that the overuse of delexical collocations in the learner corpora was due to repeated use of some combinations, so chapter 4 investigates further the nouns that form lexical patterns with the verbs. In order to delve deeper into the semantic fields formed by the noun collocates, the NS English material is supplemented by material from the BNC, using *BNCweb* to retrieve the relevant combinations. There is a thorough description of the procedure employed to retrieve collocates from the BNC, and then follow descriptions of the noun collocates of each of the six verbs, with normalized frequencies presented in tables, first outlining the results from the BNC, and then from the other corpora. The results from the BNC show that the verbs have clear semantic preferences, with ‘have’ being the most versatile verb. When it came to the learner data, however, there was underuse of delexical verb + noun constructions in most of the semantic fields identified on the basis of the BNC. This underuse is explained by a lack of variation among the learners, since “the high token frequencies in the learner corpora were a result of the repeated occurrence of a limited number of delexical verb + noun collocations” (p. 123), and this overuse of certain combinations is linked to the learners’ exposure to items that are frequent also in NS English. There was some evidence of L1 influence on the learners, reflecting both the different “world values” in the two L1 backgrounds and discourse conventions that apply in one culture or the other (p. 124). The author quite rightly points out, however, that it is often difficult to isolate one single influencing factor, and that very often there is an interplay between several factors (p. 125).

In chapter 5 the focus shifts to morphosyntactic features of the collocations under investigation, based on the fact that the grammatical aspect of collocation use can pose as many problems for learners as the

lexical aspect. Three morphosyntactic features are investigated: the morphology of the noun collocate, the type of determiner, and the use of postmodifiers (p. 128). The chapter has subsections for each of these features, and within each the overall frequency picture is presented in table and bar chart form, and then the frequencies for each of the individual verbs are presented in a table. Throughout the chapter there are examples that illustrate the phenomena discussed. The target-language (TL) norm appears to be that the collocations show a preference for “a morphosyntactic pattern where the noun collocate is identical to its simple verb stem, the determiner belongs to one of the three main types [...], and the noun collocate is modified by a prepositional-phrase postmodifier” (p. 161), although there are differences between the six verbs studied. The learners generally seemed to follow the TL norm, but there were differences between the two learner populations. As regards L1 influence, the author concludes that “grammatical features were less susceptible to L1 interference” (p. 162), and links this to the learners’ advanced proficiency level and the acquisition of certain fixed, idiom-like collocations that were used frequently in the material.

Chapter 6 discusses errors and unidiomatic usage in the learner data, and attempts to provide explanations for the types of error found in the material. It is pointed out that the present investigation, by including two learner populations with different L1 backgrounds, can supplement previous research into learners’ deviations with regard to the use of verb + noun collocations, since it can indicate problems that “are universal to all learner varieties” and not just one learner population’s problems resulting from L1 transfer (p. 166). There follows an overview of the types of errors identified in the learner material, as well as a discussion of the distribution of different types of error in the learner data. The results indicate that learners are proficient with regard to producing appropriate collocations, but that this may be linked to a certain caution on their part, since they rely on highly frequent patterns to a large extent. Some verb + noun collocations cause more problems than others.

As regards the two groups of learners, they share all the problems identified, although overall the Swedish learners produced fewer errors than the Chinese learners. The Swedish learners had a higher proportion of errors related to lexical choice (e.g. by confusing the verbs ‘make’ and

'do'), while the Chinese learners made more grammatical mistakes, typically by leaving out articles (p. 198).

Summary and conclusions are presented in chapter 7. The summary is structured around the research questions, with each question being listed followed by a discussion that summarizes the results and thereby answers the question. The first question was concerned with the frequency of delexical verb + noun collocations in the learner material. The results indicated overuse of such collocations by both learner groups, and this overuse was caused by "the repeated occurrence of a limited set of collocations" (p. 202). This is interpreted as being a developmental interlanguage feature.

The second question dealt with the extent to which the learners resemble or differ from native speakers of English and L1 Swedish and Chinese in their lexical choices and the morphosyntactic features of the collocations. It was found that the semantic fields of the noun collocates in NS English were more varied than those found in the learner data, and that the learners may be influenced by the cultural conventions of their respective countries of origin. However, "[m]any of the frequent collocations in the [target language] were found in both learner corpora, suggesting a positive effect of [target-language] frequency on the [interlanguage]" (p. 202). As regards syntactic patterns, the learners performed very similarly to the target-language writers, possibly due again to the reliance on a small number of high-frequency collocations by the learners.

The proportion of errors was the focus of the third research question, and here the results indicated that the learners faced similar problems, albeit with the differences outlined in the discussion of chapter 6 above.

As regards factors influencing the learners, the author states that the learner populations were more similar than different, despite their very different L1 backgrounds, and that "only partial support" has been identified for L1 influence (p. 204-205). The study also revealed "a subtle interplay between interlingual and intralingual factors (e.g. essay topic, the 'teddy-bear' principle, the complexities of the [target-language] constructions) that may have influenced the learners' use of delexical verb + noun constructions" (p. 205).

The chapter ends with a discussion of pedagogical implications and suggestions for future research. With respect to the former, the author suggests that delexical verb + noun collocations remain problematic even

for advanced learners, and that the findings of this study indicate areas that need to be emphasized in language teaching, by means of explicit instruction and an increased focus on “the essentially phraseological or formulaic nature of ordinary language use” (p. 207). With respect to the latter, i.e. suggestions for further research, it is suggested that, given the importance of the input received by learners for the acquisition and use of delexical verb + noun collocations, it would be fruitful to conduct a detailed study of the treatment of phraseology in textbooks and other teaching materials (p. 207). Supplementing corpus data with elicitation data is also suggested, as are psycholinguistic experiments in order to further our understanding of how language is processed in learners’ minds.

Wang’s book is a very thorough account of the structure and use of delexical verb + noun collocations in English produced by Swedish and Chinese learners. The main strengths of the book lie in the very clearly defined search procedures and analytical frameworks employed, as well as in the design of the study: the inclusion of L1 corpora of Chinese and Swedish is highly commendable, in that it allows for empirical conclusions to be drawn about potential L1 influence in addition to an investigation of developmental factors and interlanguage universals.

There are some limitations, however, and these are chiefly to do with the quantitative methods. There is only a very brief discussion of the statistical procedures used, where it is stated that log-likelihood is used as the statistical measure for reasons outlined in another publication (p. 35). In later chapters, however, it is clear from the statistical output presented that chi-square calculations have been performed instead (see e.g. p. 132), but the reason for this change is not mentioned. Nor is there an account of how the familywise error rate (cf. e.g. Field et al 2012: 428-431) has been adjusted for, so it is not clear whether the tests performed are actually valid. Finally, the author presents results in the form of tables and bar charts containing normalized frequencies per 100,000 words (see e.g. Table 3.1 and Figure 3.1) or percentages (see e.g. Table 5.1 and Figure 5.1). This does not take corpus-internal dispersion into account, and especially with low-frequency items it disguises the extent to which individual writers potentially skew the data. One may also ask whether a verb collocation that occurs 22 times in 230 essays can really be called a “frequent noun collocation” of ‘have’ used by Chinese learners, as is done with the noun ‘right’ in Table 4.3 (raw

frequency found in Table 1 in Appendix II). This is just a single example, but it illustrates a phenomenon that occurs multiple times in the book. It would surely be more appropriate to say that Chinese learners do not use any noun frequently in combination with 'have', since at most this collocation is used by 9.56% of the Chinese learners (and that applies only if those learners use it one time each). A more interesting perspective would perhaps be to discuss what the majority of, in this instance, Chinese learners do instead.

*Sylvi Rørvik*

Inland Norway University of Applied Sciences

*References*

Field, Andy, Jeremy Miles and Zoë Field. 2012. *Discovering statistics using R*. London, Thousand Oaks, New Delhi, Singapore: SAGE publications.