

The Brazilian Lexicographic Road to Bilingual Verbal Collocations

Stella Esther Ortweiler Tagnin

University of São Paulo
Av. Prof. Luciano Gualberto, 403
05508-900 – São Paulo – SP
Brazil
e-mail:seotagni@usp.Br

Abstract

Within the realm of collocations, a highly neglected category in terms of lexicographic material available is the category of verbal collocations. With the exception of a few monolingual general collocational dictionaries and even fewer bilingual similar dictionaries, there is only one specific bilingual dictionary of verbal collocations, the *Russian-English Dictionary of Verbal Collocations* [Benson & Benson 1993]. As it is highly desirable that this type of dictionary exists for every pair of languages, we have engaged in compiling a bi-directional English-Portuguese dictionary of verbal collocations in Brazil. The process of compilation and the problems arising thereof are briefly discussed in this paper.

Introduction

Now that the study of lexis has come to the fore and is beginning to receive due attention, especially for the role it plays in nativelike production, the lack of lexicographic material in the area of phraseology is ever more evident. With the exception of the monolingual *BBJ Combinatory Dictionary* [Benson et al. 1986] and the *LTP Dictionary of Selected English Collocations* [Hill & Lewis 1997] and a couple of bilingual dictionaries of collocations (in Chinese and Japanese), there are no other such dictionaries. When it comes to a special category like verbal collocations the situation is even more dramatic: there is only the *Russian-English Dictionary of Verbal Collocations* [Benson & Benson 1993], when it would be highly desirable to have these reference sources for as many pairs of languages as possible. For this reason we have engaged in compiling a bilingual English-Portuguese/Portuguese-English dictionary of verbal collocations. In this paper I will discuss the problems encountered and the decisions made during its process.

Collocations and Nativelike Production

It is a known fact that a person's fluency is closely related to his/her knowledge of the "building blocks" that make up a large part of the language. Pawley & Syder [1983] have pointed out how intimately related phraseological competence and nativelike production are. Fillmore [1979] coined the term "innocent speaker" for anyone who is not aware of what he calls the "conventionality" (now commonly known as phraseological) aspect in language. Although this element is usually brought up in relation to learners of a foreign language, native speakers also show deficiencies in this aspect. Azevedo [s/d], in a study of pre-university entrance examinations of Brazilian students, discusses various collocational inaccuracies, highlighting that "verbal collocations are undoubtedly the most problematic category."

Collocations and Dictionaries

Collocations are mainly a production problem and therefore require a specific lexicographic format. Because the consultant does not know the *verb* that regularly cooccurs with a certain noun, the entry should be made by the noun [Hausmann 1988], which is not how most general language dictionaries list verbal collocations – if they do at all. Lexicographic deficiencies in this aspect have been revealed by various scholars [Bahns 1987; Heid & Freibott 1991; van der Wouden 1992], mostly in relation to English dictionaries. The same is true for Brazilian dictionaries, both monolingual and bilingual [Tagnin 1999], though it may not hold for other language pairs.

In view of the above, a bilingual dictionary of verbal collocations seems to be a highly desirable solution. Strangely enough, though, there is, to our knowledge, only the *Russian-English Dictionary of Verbal Collocations* [1993] by Benson & Benson.

To help fill this gap we would like to present the *Bilingual English-Portuguese/Portuguese-English Dictionary of Verbal Collocations*.

Bilingual English-Portuguese/Portuguese-English Dictionary of Verbal Collocations.

This dictionary is mainly aimed at a text producer, which ranges from a learner of English or Portuguese, to a translator into either language, or a native writer in either language. Although it is bi-directional, it can also be consulted as a monolingual dictionary, as will become clear to the reader in the course of this article. It currently features over 4,000 collocations in each language.

The initial stages of the project

We began collecting our data in a pre-corpora era (at least in Brazil) and, as Mackin stated

The compiler of a dictionary of collocations has three main sources open to him: first, other dictionaries; second, his own 'competence'; and third, occurrences met with in the course of reading and listening to the spoken word on radio, on television, in conversation, at lectures, at the cinema, and so on. [1978:152]

Intuitively, we resorted to all three sources. First, - this was in 1988 - students of a Diploma course in Translation at the University of São Paulo were assigned one or more letters of the alphabet and instructed to consult, in that order, a monolingual English dictionary, a bilingual English-Portuguese dictionary and, finally, a monolingual Portuguese dictionary.

They should be on the lookout for verbal collocations listed as entries, subentries, part of a definition or of an example. Occurrences were jotted down on cards for group discussion until all achieved a clearer notion of what a verbal collocation was for quite often they confused them with idiomatic expressions. This is where 'competence' came into the picture.

The next step was collecting authentic examples, mainly from written sources: books, newspapers, magazines, etc. At one point undergraduate students were engaged in the

project and required to present a certain number of English verbal collocations with authentic examples as their final paper for the semester.

In the early 90s, a corpus of Brazilian Spoken Language, the NURC, became available in electronic format [Castilho et al. 1986-1988] and was consulted for Portuguese verbal collocations.

It was only in 1990 that the data collected was fed into a computer with the help of the MC4 – Microcezeau software, a dBase database, based on DOS, installed in a 286 computer, with a 512 MB hard disk and 16 MB RAM, without a 3 ½" diskette drive. Because the program did not run under Windows, all data had to be inserted manually. In other words, although there was a large amount of data that had been submitted in diskette, it was not possible to “cut and paste” this material into the database.

Only a couple of years ago was an Access-based program developed which now allows insertion of data under Windows via a “cut and paste” procedure. This has visibly sped up the process. This database is reversible, producing output with either language as the source language.

Access to Electronic Corpora

Let us return to the history of the project. In 1991, during a visit to Brazil, Tim Johns, from the COBUILD project at the University of Birmingham, offered us a working copy of his concordancer. Even though it was not a full-fledged search engine, it was extremely valuable at that point to complete and confirm certain data.

Our first contact with a large corpus occurred in 1995, at the Universities of Gotemburg and Lund, when we were able to use the Brown Corpus, thanks to Göran Kjellmer and Bengt Altenberg. With a list of collocations from our corpus for which we had no examples, we were able to complete, in a few days, what would have taken us months in Brazil, where corpora were not available.

In the meantime, we have acquired a few corpora on CD-ROM, including the Brown Corpus and complete versions of one of our daily local newspapers, *Folha de São Paulo*, for the years 1994-1999. The latter's built-in search engine is Folio, which does not produce concordances but lists each occurrence in a larger context separately, which makes the search somewhat cumbersome. No statistic method has been used to determine their collocational status. Rather, we have relied on recurrence, that is, we have set a minimum of three authentic examples for a combination to be considered a collocation. But we have also relied on intuition because certain combinations – though not frequent – are collocations all the same. Such is the case of *dirimir uma dúvida*, for example: in the above *Folha* corpus for 1999, consisting of 100,000 texts, *dirimir* only occurs in 24, out of which it occurs 13 times with *dúvida(s)*. On the other hand, the same corpus for the year 1998 features *dirimir* in 28 texts, but it only cooccurs with *dúvida(s)* in 6.

Syntactic Analysis

A syntactic analysis produced the following basic patterns [Tagnin 1999]

- a. verb + **noun** _[object] : *take advice - aceitar (um) conselho*
- b. **noun** _[subject] + verb: *a river flows - um rio corre*
a film/play opens - um filme/uma peça estréia
- c. verb _[copula] + **adjective** : *fall ill - ficar doente (adoecer)*
fall asleep - cair no sono (adormecer)
run dry - ficar seco (secar)
come true- tornar-se realidade (realizar-se)
- d. verb _[movement] + **adverb**: *climb aboard - subir a bordo*
go ashore - descer a terra ,
lead astray - levar para o mau caminho
- e. verb _[static] + **advérbio**: *sit/stand astride - sentar/ficar de pé de pernas abertas*

In English there are also verbal collocations constituted of a nonexistent category in Portuguese, the *phrasal verb*, such as, *put on an act* and *keep up appearances*.

This classification proved to be relevant to define our basic unit of analysis. However, a more detailed analysis revealed the occurrence of prepositions in certain structures such as *take into account/levar em conta*, so that our unit

- is not restricted to monolexemic verbs, thus including phrasal verbs,
- includes prepositions so as to provide information about its colligation.

Semantic Analysis

Semantic analysis was carried out on two levels: between verbal collocations to account for hyponymy and superordination, and between subject and predicate. We initially resorted to Mel'cuk's Lexical Functions but found them not to be user-friendly. Van der Wouden [1992:454] also questions "whether an ordinary user will be able to fruitfully use such a dictionary?" In other words, a consultant would have to be acquainted with approximately 50 such functions before being able to understand an entry. Contrariwise, the entry as devised for this dictionary, will allow the consultant to infer the difference in meaning between collocations occurring with the same base – i.e. *draw attention* and *pay attention*, for instance – from various types of information provided in the entry: the examples the source language, the equivalents in the target language – where the same distinctions should prevail – and the examples in the target language.

The overall structure of the dictionary

The dictionary consists of two parts: English-Portuguese and Portuguese-English. All collocations are listed under their base. These are followed by the information of their grammatical category (*noun, adjective or adverb*). One of the problems encountered was to decide under which noun to list a collocational range like *take a pill/an aspirin/a tablet/medicine*. A possible solution would be to list it under the superordinate, *medicine* in this case. Though economical, it did not seem to be the best solution: how could one be sure the user would arrive at *medicine* as the superordinate. Why not *remedy*, for example? In

view of this, it was decided to list each occurrence, adding the information that the noun is either a superordinate or a hyponym. A superordinate is indicated by the abbreviation *gen.* (for “general”) and a hyponym by *part.* (for “particular”). Linguistic terminology has been avoided to make the dictionary more user-friendly: **medicine** *n. gen.* Or **aspirin** *n. part. [medicine]*.

Whenever possible, the equivalent in the target language will also be a collocation. When this is not the case, a functional translation will be offered:

take action → *tomar uma atitude* (a collocation in both languages)
tell apart → *distinguir* (a functional translation in Portuguese)

The examples provided are all authentic, extracted from various written and spoken sources, as well as corpora. Reference are given in parentheses (see below).

Another problem we encountered was related to selecting the collocations to be included. It has been argued that the teaching of collocations should concentrate on those which are not easily “transferred” into the target language. However, if an isomorphic collocation is excluded it might lead the consultant to believe that it does not occur in the target language at all. Therefore, again aiming at a user-friendly source of reference, we have included both isomorphic and nonisomorphic collocations.

The microstructure of the dictionary

To ensure a systematic treatment of collocations in our dictionary they are, as stated above, inserted by the base. The meaning of the base is given in square brackets only when it is polissemic: **account** *n.* [explanation].

Synonymous collocates are listed together separated by a slash. Register information is added when adequate: *form.* stands for “formal”, *inf.* for “informal”, for instance. Duly identified authentic examples follow:

advice, follow/take/ (*form.*) **heed sb's;**
If you want to change your emotional style, you might want to follow the advice given by a man who came to one of my workshops. (Siegel)

Next the equivalent collocation, if there is one in the target language, is given. In case there is none, a pragmatic translation is offered and marked off by a symbol:

aground, run
The ship has simply run aground and broken up on the reef.
 → **encalhar**

In case two or more equivalent collocations are possible, they are listed in alphabetical order. All are followed by examples, which are not translations of the source language examples, rather they are authentic instances of usage of the target language collocation:

aid, cut off
The Gulf states punished Jordan for its pro-Saddam tilt by cutting off foreign aid...
 (Coll.)
ajuda, cortar uma
Nenhum governo pode cortar sua ajuda a países pobres... (FSP95)
auxílio, cortar um
... EUA para não cortar seu auxílio aos países mais pobres. (FSP95)

As to the article that may precede the base, there are four possible situations: a. the base may take both a definite or an indefinite article – this is indicated by an indefinite article: **alibi, support an**; b. the base may or may not take an article – this is indicated by an indefinite article in parentheses: **allowance, make (an) ~for**; c. the base takes necessarily a definite article: **air, clear the**; d. the base takes no article: **allegiance, give/pledge/swear**.

Because this has been a long process, not always carried out in a systematic manner, each entry is currently being revised to conform to the above norms. This last phase should be completed by the end of this year.

References

- [Azevedo s/d] Azevedo, Laís Furquim (s/d). Deslizes lingüísticos em colocações convencionais, MS.
- [Bahns 1987] Bahns, Jens, 1987. Kollokationen in englischen Wörterbüchern, in: Diller, Hans-Jürgen et al. (eds.). *Wortschatzarbeit, anglistik & englischunterricht*, Heidelberg, Carl Winter Universitätsverlag, 87-104.
- [Benson & Benson 1993] Benson, Morton & Evelyn Benson, 1993. *Russian-English Dictionary of Verbal Collocations*, Amsterdam/Philadelphia, John Benjamins.
- [Benson et al. 1986] Benson, Morton, Evelyn Benson & Robert Ilson, 1997 [1986]. *The BBI Combinatory Dictionary of English - A Guide to Word Combinations*, Amsterdam/Philadelphia, John Benjamins.
- [Castilho et al. 1986-1988] Castilho, Ataliba Teixeira, Dino Preti & Hudinilson Urbano (orgs.), 1986-1988. *A Linguagem Falada Culta na Cidade de São Paulo vols. 1-3*, São Paulo, T.A. Queiroz/FAPESP, put in electronic format by Zilda Maria Zapparoli Castro Melo and Leland McCleary, FFLCH/USP.
- [Fillmore 1979] Fillmore, Charles J., 1979. Innocence: a second idealization for linguistics, *Berkeley Linguistic Society* 5, 63-76.
- [Hausmann 1988] Hausmann, Franz Josef, 1988. Grundprobleme des zweisprachigen Wörterbuchs, in Hyldgaard-Jensen, Karl & Arne Zettersten (eds.). *Symposium on Lexicography III, Proceedings of the Third International Symposium on Lexicography*, May 14-16, 1986, at the University of Copenhagen, Tübingen: Niemeyer, 137-154.
- [Heid & Freibott 1991] Heid, Ulrich & Gerhardt Freibott, 1991. Collocations dans une base de données terminologique et lexicale, *Meta XXXVI:1*, 77-91.
- [Hill & Lewis 1997] Hill, Jimmie & Michael Lewis (eds.), 1997. *LTP Dictionary of Selected Collocations*, Hove, England: Language Teaching Publications.
- [Mackin 1978] Mackin, Ronald, 1978. On collocations: 'words shall be known by the company they keep', in: Strevens, P. (ed.). *In Honour of A. S. Hornby*, Oxford: Oxford University Press.
- [Mel'cuk et al. 1984] Mel'cuk, I., N. Arbatchewsky-Jumarie, L. Elnitsky, L. Iordanskaja & A. Lessard, 1984. *Dictionnaire Explicatif et Combinatoire du Français Contemporain: Recherches Lexico-Sémantiques I*, Montreal, Les Presses de l'Université de Montréal.
- [Pawley & Syder 1983] Pawley, A. & F.H. Syder, 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency, in: Richards, Jack C. & Richard W. Schmidt (eds.), *Language and Communication*, London/New York, Longman, 191-225.
- [Tagnin 1999] Tagnin, Stella E. O., 1999. *Convencionalidade e Produção de Texto: Um Dicionário de Colocações Verbais Inglês/Português, Português/Inglês*, Post-Doctoral thesis, Department of Modern Languages, University of São Paulo, São Paulo.
- [van der Wouden 1992] Van der Wouden, Ton, 1992. Prolegomena to a multilingual description of collocations, in Tommola, Hanny, Krista Varantola, Tarja Salmi-Tolonen & Jürgen Schopp (eds.) *Euralex '92 Proceedings, vol II*, University of Tampere, Finland, 449-456.