

FORBONNAIS'S FIRST PORTUGUESE TRANSLATION: BUILDING A LEXICOGRAPHIC CORPUS

ESPERANÇA CARDEIRA, , JOÃO PAULO SILVESTRE, ALINA VILLALVA
UNIVERSITY OF LISBON - CENTRO DE LINGUÍSTICA
LISBON, PORTUGAL

ABSTRACT

O projecto EE-T disponibiliza um fundo textual diacrónico, de grande interesse para os estudos lexicais e lexicográficos. Neste trabalho discutimos a possibilidade de os textos resultantes de traduções poderem ser integrados num corpus documental multilingue, para documentação de dicionários terminológicos e de dicionários bilingues. Procuramos também avaliar a possibilidade de textos obtidos a partir de reconhecimento automático de caracteres serem considerados testemunhos válidos de domínios terminológicos específicos. A edição da primeira tradução portuguesa de Forbonnais (Elementos do Commercio, 1766) servirá de estudo de caso.

1. A documentação em terminologia

Corpus lexicográfico e corpus documental são conceitos distintos. Muitos dicionários terminológicos especializados enunciam um amplo conjunto de fontes, que sustentam a actualização da lista de termos. Todavia, em boa parte dos casos esse corpus apenas serve de suporte para as citações e autorizações da terminologia previamente conhecida ou esperada. Consideremos a informação que se encontra no prólogo de duas edições do Routledge Dictionary of Economics de Donald Rutherford

«Economics, the Queen of the Social Sciences, has now established itself as a major subject in dialogue with the physical sciences, law and the arts.»

To produce a dictionary of this kind, I started with an assortment of basic textbooks and many current newspapers and journals. I soon discovered that about a thousand concepts are common to all the textbooks, for example notions of cost, economic systems and banking. From general textbooks I moved to a perusal of specialist books on the diverse divisions of the subject. The areas of economics encompassed obviously have to reflect current concerns; many environmental concepts are included and the 'male' character of many economics works has been partially avoided by including biographies of several leading female economists. Newspapers and journals provide a modern guide to current economic discourse. There is no foreseeable end to the creation of economic neologisms – major events such as the deregulation of financial markets and the political developments in Eastern Europe, which have changed the nature of many economies, have produced an expansion of new terms. Some terminology is ephemeral but many words that start as slang, such as 'yuppy', have a surprising longevity. I have taken the optimistic view that numerous catchwords and catchphrases will render linguistic service for many years.

1st. ed. 1992

Ten years have elapsed since the first edition of this Dictionary. The vocabulary of economics in the broadest sense has considerably grown. Many neologisms have sprung from continued changes in national economies, not least the innovations in financial markets and growing concerns about the environment. Institutional changes, for example, the coming of the World Trade Organisation, and new interests in economic thought, not least through the further awards of Nobel Prizes for Economics, have inspired new entries. Inevitably some terms in the first edition have not been as durable as others and recommended reading needed revision.

Extensive reading of economics journals and monographs, as well as newspapers, has produced over a thousand new entries.

2nd ed. 2002

Usar um corpus de citações condiciona o resultado da pesquisa lexicográfica de terminologias.

Until about 1980, the main form of empirical language data available to lexicographers was the citation. A citation is a short extract from a text which provides evidence for a word, phrase, usage, or meaning in authentic use.

Although instances of usage are authentic, there is a big subjective element in their selection. Human readers tend to notice what is remarkable and ignore what is typical, and this creates a prejudice towards the novel or idiosyncratic usages which inevitably catch the reader's eye.

2. Características de um corpus lexicográfico

Objective evidence of language in use is a fundamental prerequisite for a reliable terminology dictionary. Traditionally, such evidence was found in collections of citations, but these have their limitations. If the dictionary's function is, as we have argued, to focus on 'normal' language events, it follows that you need very large volumes of data: normal language events are those which are recurrent, which can be observed to take place frequently and in different types of text.

So we can only confidently distinguish what is conventional from what is idiosyncratic if we have plenty of data at our disposal.

So most corpora will comprise a subset (usually a very small subset) of all of the communicative events of the language under investigation. It must, in other words, be a sample. To create a sample that fairly reflects the wider population, you need clear selection criteria, and these will be determined by your corpus's intended function. People use corpora for all sorts of purposes, many of them highly specialized.

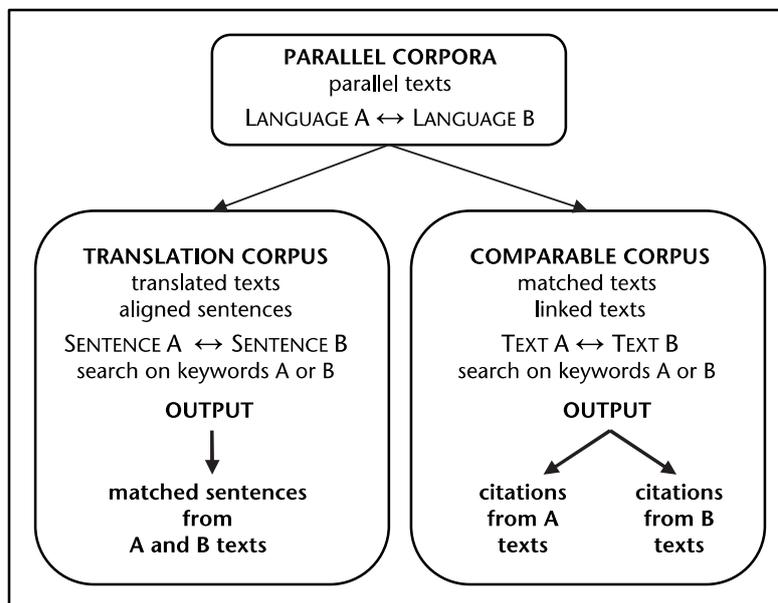
Selecting texts on the basis of their 'quality', and excluding those which fail this test, is fundamentally at odds with the descriptive (as opposed to prescriptive) ethos of corpus linguistics. Who is to judge which texts are 'good', and on what basis? The whole point of using corpora is to avoid pre-judging the data and choosing texts because you approve of them in some way.

3. Características específicas de um corpus bilingue

There is an important difference between translating for terminology dictionaries and the more familiar discourse- or text-translating. In the case of extended discourse, as opposed to dictionary entries, a good translation will produce language so idiomatic and natural sounding that the reader may not be aware that it is a translation. It is well known, however, that on analysis the original texts (in dictionary terms, the source language, or SL) and the translation (the target language, or TL) rarely align perfectly, in that the sense of one individual word in the original is not exactly reflected in any corresponding TL word. Rather, the sense of a longer stretch of SL text is rendered in a corresponding piece of TL text.

The term parallel corpus denotes a set of corpora (two in a bilingual parallel corpus, more in a multilingual version) in which the texts in Language A correspond in some way to those in Language B (and perhaps C and D and so on).

In a translation corpus, the two corpora consist of translated texts, which means of course that only 50 per cent of the texts are originals, the rest being translations. Nonetheless, a translation corpus is a rich source of equivalence material and easy to use, as software exists to align pairs of sentences, one from each language corpus.



A comparable corpus is made up of two individual language corpora, selected on the basis of at least one shared parameter, usually the subject matter, together with possibly other properties shared by the texts, such as the date and/or the medium (books, newspapers, conversations, etc.).

This type of corpus provides excellent material for the translator because all the texts are original, and no translation is involved. For bilingual lexicographers it could be a rich source of inspiration, but, because the matched corpora can only be searched individually, the output is not economic to use in a serious dictionary project.

4. Constituição de um corpus lexical a partir da edição de *Elementos do Commercio*

Abordam-se os seguintes aspectos:

Características materiais do texto e problemas para a aplicação de reconhecimento automático de caracteres (apresentação de resultados sem intervenção);

Descrição de intervenções de edição automática do texto, a partir do conhecimento das características morfológicas das unidades lexicais;

Elaboração de índices de frequência das unidades reconhecidas e avaliação dos domínios lexicais mais representados.



1 paiz.

DO COMMERCIO. 5-1

paiz. Quatro meios ha para iilo for-; tir effeito : a concorrência , a economia do trabalho dos homens , o módico preço das difpezas da exportação , e o preço baixo do juro do dinheiro.

A concurrencia prodús a abundância , e esta o bom preço dos vi- - veres , das matérias primeiras, dos artiftas , e, do dinheiro. A concorrência he hum dos mais importantes princípios do Commercio , e huma parte confiideravel da fua liberdade. Tudo o que a conítrange , ou a altera nestes quatro pontos , he prejudicial para o Eftado diametralmente oppoito ao feu objecto , que he a felicidade , e a fubíiftencia commoda do maior numero de homens.

A economia do trabalho dos ho mens conííte em fupprir pelo das máquinas , e dos animaes , quando* pode fazer-fe com menos cufto : o que faz multiplicar a povoação em lugar de a deftruir. Efte ultimo er- vTO fe confervou muito tempo nospai- zes, que não fe occupavaõ mais que «o Commercio interior. E com ef- D ii feita



e se elle os vender mais baratos ,
terá a preferencia até no seu mesmo
. »V-i paiz.

DO COMMERCIO. \$1

paiz. Quatro meios ha para isto for-; tir effeito : a concorrência , a
economia do trabalho dos homens , o módico preço das dispezas
da exportação , e o preço baixo do juro do dinheiro.

A concurrencia prodús a abundância , e esta o bom preço dos
víveres , daf matérias primeiras, dos artistas , e, do dinheiro. A
concorrência he hum dos mais importantes princípios do
Commercio , e huma parte consideravel da sua liberdade. Tudo o
que a conílrange , ou a altera nestes quatro pontos , he prejudicial
para o Estado diametralmente oppoíó ao seu objecto , que he a
felicidade , e a subsistencia commoda do maior numero de
homens.

A economia do trabalho dos ho
mens consiste em supprir pelo das
máquinas , e dos animaes , quando
pode fazer-se com menos custo : o
que faz multiplicar a povoação era
lugar de a destruir. Este ultimo ertro se confervou muito tempo
nos paizes, que não se occupavaõ mais que
«o Commercio interior. E com efD li feito,

* ELEMENTOS

feito se o Commercio exterior formediocre , o objedk) geral mg
terá