What Can Metaphor Tell us about the Language of Translation?

Sara Piccioni*

Università “G. D’Annunzio” di Chieti-Pescara, Viale Pindaro, 42 – Pescara 65127 Italy

Abstract

This paper illustrates an exploratory study aimed at devising a methodology for the analysis of the language of translations through a comparison of metaphor use in original and translated texts. It uses a pilot monolingual comparable corpus of corporate sustainability reports made up of 2 sections: a subcorpus of Spanish originals and a subcorpus of translations from English into Spanish. VERB-NOUN metaphors are analyzed to compare collocation variety, typical collocations and degree of metaphorical conventionality of the VERB-NOUN pairs in original and translated texts. Results suggest that metaphors in translated texts show both a tendency to normalization and a preference for unconventional uses arising from original text expressions “shining through” in the translations.

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of CILC2013.

Keywords: Metaphor; Translation Universals; Monolingual Comparable Corpus; Spanish; Sustainability Reports.

1. Introduction

In recent decades, Translation Studies have witnessed a move towards descriptive approaches focusing on the analysis of translated texts in order to identify regularities and “norms” characterizing the language of translations. The underlying assumption is that translations display linguistic features that result from the translation process itself, and that differentiate translated language from that of original texts (Baker, 1993).

This paper illustrates an exploratory study aimed at devising a methodology for the analysis of the language of translations through a comparison of metaphor use in original and translated texts. It is hypothesized that metaphor provides a privileged standpoint to explore the identifying traits of translated Spanish and, in particular, to test the
validity of the normalization/conventionalization hypothesis, i.e. the idea that translations tend “to exaggerate features of the target language and to conform to its typical patterns” (Baker, 1996).

The analysis focuses on a pilot monolingual comparable corpus of corporate sustainability reports made up of 2 sections: a subcorpus of Spanish originals and a subcorpus of translations from English into Spanish. VERB-NOUN metaphors are analyzed to compare collocation variety, typical collocations and degree of metaphorical conventionality of the VERB-NOUN pairs in original and translated texts. Results suggest that metaphorical use in translated texts shows both a tendency to normalization and a preference for unconventional uses arising from original text expressions “shining through” in the translations.

After a brief overview of the Translation Universal hypotheses (TU) which provides the theoretical framework for the study (Section 1.1), the article describes the corpus used for analysis and the steps followed for its compilation and annotation (Section 2). In Section 3, the methodology used for data extraction and the results of the analysis are illustrated. The final Discussion (Section 4) assesses the results against the TU framework and extrapolates methodological indications for future work.

1.1. The Translation Universal hypotheses

The Translation Universal hypotheses identify a body of work within Corpus-Based Translation Studies which has focused on the identification of the linguistic features that make translated texts different from original texts. The awareness that “translations are produced in a different context, under different constraints and pressures, and respond to different influences and motivations” (Dayrell, 2007: 375) has led translation scholars to consider these texts in their own terms as representative of a “third code” (Frawley, 1984), different from the language used in original texts.

The main TU discussed in the literature can be summarized as follows:

- Simplification, defined as the “tendency to simplify the language in translation” (Baker, 1996). (Laviosa, 1998) identifies this tendency in English translations displaying lower lexical density, limited vocabulary, greater proportion of high-frequency words, and more frequent repetition of frequent words than original texts.
- Explicitation, i.e. the tendency “to spell things out in translation” (Baker, 1996). This has been mainly investigated at syntactic/textual level, e.g. by (Olohan & Baker, 2000) who find a higher proportion of expressed (optional) reporting that in translations than in original texts.
- Normalization/Conventionalization, i.e. the “tendency to exaggerate features of the target language and to conform to its typical patterns” (Baker, 1996). This can be seen in the greater use (in translation) of idioms or of typical grammatical structures, elimination of sociolinguistic features, or exaggeration of typical features of a given genre. Kenny (2001) observes this tendency at lexical level in the “neutralization” of atypical collocations.
- “Shining through” of the original language in the target language. This is seen when “in a translation into a given target language (TL), the translation may be oriented more towards the source language (SL)” (Teich, 2003).

For the purpose of this study the last two notions are particularly relevant. The frequency of metaphors and their degree of conventionality in original and translated texts can be useful parameters in identifying a text’s degree of idiomaticity, thus allowing testing of the normalization hypothesis. Conversely, “shining-through” effects could result from the literal translation of source-text metaphors which do not have an immediate or conventional equivalent in the target language. This aspect has already been observed in specialized texts where the literal translation of source-text metaphors has been shown to contribute to the introduction of neologisms in the target language (Samaniego Fernández, Velasco Sacristán & Fuertes Olivera, 2005).

2. Corpus design, compilation, and annotation

The study here proposed focuses on a corpus of corporate sustainability reports (also known as Corporate Social Responsibility Reports, CSR reports henceforth). These are reports issued every year by private companies and organizations to account for their “economic, environmental, social and governance performance” (Global Reporting Initiative). While allowing for a degree of flexibility, the texts can be ascribed to a single, well-defined
genre whose content and form respond to specific constraints set by a number of standardization initiatives promoted by organizations such as the Organization for Economic Cooperation and Development, the United Nations Global Compact, and ISO. By restricting the corpus to a single genre the study looks to reduce the number of variables involved in the interpretation of results; furthermore, the choice of a non-literary, yet not overly specialized genre, allows observation of both conventional (e.g. idioms) and creative metaphors (e.g. those arising from the creation of neologisms).

In its final configuration the corpus will display two main components. First, a Spanish monolingual comparable corpus will comprise a subcorpus of original CSR reports (ES-OR henceforth) and a subcorpus of CSR reports translated from English (ES-TR henceforth). Second, a bilingual parallel corpus will allow alignment and comparison of Spanish translated reports (ES-TR) and their English originals (EN-OR henceforth). As in previous work within Descriptive Translation Studies (e.g. Bernardini, 2007; Hansen-Schirra, 2011), the monolingual comparable corpus will be used to compare the language of original and translated texts in Spanish (ES-OR vs. ES-TR), while the parallel corpus will be used to trace specific Spanish target-text instances (ES-TR) back to their English originals (EN-OR).

For the purposes of this exploratory study a limited portion of the Spanish monolingual comparable component of the corpus was compiled. This was done by downloading original and translated CSR reports from corporate websites, selecting reports published between 1999 and 2012. While the ES-OR subcorpus only includes full reports, the ES-TR includes both full reports and summaries: this is due to a common practice whereby some companies only translate extracts or summaries of the full reports in languages other than English. As a result, ES-TR includes a higher number of shorter texts, which could partly impair quantitative comparability of the two subcorpora. A further consideration to keep in mind in terms of corpus compilation regards the sectors within which each reporting company operates and the influence this could have on the language and metaphors used in the texts. In an effort to reduce variability, this study focuses on reports by companies operating in either the banking or energy sectors. In the future a bigger, and more varied, corpus will enhance comparability of the original and translated components of the corpus.

In order to make the corpus searchable, the reports downloaded in .pdf were converted into .txt and edited by eliminating (among other things) lists of contents, page headers and footers, the content of tables, and corporate charts containing long lists of English words and acronyms. Each subcorpus was then lemmatized and POS-tagged using the TreeTagger (Schmid, 1994) and indexed with the Corpus Work Bench (CWB) to allow interrogation with the Corpus Query Processor (CQP; Christ, 1994). There resulted two pilot subcorpora of original CSR reports (ES-OR, 51,431 tokens) and, respectively, translated reports (ES-TR, 54,176 tokens).

3. Data selection and analysis

This exploratory study focuses on metaphors that arise from a semantic clash between a verb and the noun(s) expressing its direct object (e.g., lanzar un programa). The VERB-NOUN pairs were selected by identifying the key-nouns of the subcorpora. The log-likelihood measure (Dunning 1994) and a reference corpus of newspaper articles were used to identify the words that show a stronger statistical association with each corpus when compared to the reference corpus, thus obtaining a list of the 50 most “typical” words for both ES-OR and ES-TR. Only the nouns that appear in both lists (i.e., nouns that are “typical” of both subcorpora) were selected for analysis, obtaining a list of 16 key-nouns.

The VERB-NOUN pairs were selected by extracting the verbs appearing 3 words to the left of the 16 key-nouns found in both ES-OR and ES-TR, discarding the pairs with frequency lower than 2 and those where the noun is not the verb’s direct object. Following this selection, a total of 85 VERB-NOUN pairs were considered for analysis.

The analysis focused on 2 main aspects: collocation variety and typical collocations in ES-OR vs. ES-TR; and degree of metaphorical conventionality of the VERB-NOUN pairs in ES-OR vs. ES-TR.
3.1. Collocation variety and typical collocations in original and translated texts

While not strictly related to metaphorical use, collocation variation and “typicality” were considered in order to verify whether previous hypotheses hold true in our corpora and to allow fine-tuning of the methodology used in this pilot study for future research. Collocation variety of the VERB-NOUN pairs in original and translated texts was studied by comparing the number of verbal collocates the nouns displayed in each corpus. To account for variation in the frequency of each noun in the two corpora, the frequency of the noun per 1,000 words of the corpus was divided by the number of its verbal collocates \((fq(N)/no. \text{ of collocates})\). A lower ratio indicates a higher number of verbal collocates, and therefore greater lexical variety. As shown in Table 1, after discarding the nouns with no verbal collocates in one of the two corpora, in 9 cases out of 13 translated texts display a higher number of verbal collocates than the original texts. While the figures seem to point to a higher degree of lexical variety in the translated texts (which contradicts previous findings, e.g. Dayrell 2007), the results could be indicative of greater syntactic complexity in the original texts. The 3-word span established for selecting VERB-NOUN pairs may not be big enough to spot all the nouns’ verbal collocates, which in the future can be corrected by setting a wider span for collocate extraction.

| NOUN    | OR-ES | TR-ES |
|---------|-------|-------|
| negocio | 0.47  | 1.04  |
| emisión | 0.33  | 0.34  |
| grupo   | 0.9   | --    |
| informe | 0.87  | 1.59  |
| programa| 0.63  | 0.26  |
| riesgo  | 1.01  | 0.31  |
| proyecto| 0.72  | 0.36  |
| desarrollo| 1.02 | 0.7   |
| resultado| 0.85 | 0.37  |
| objetivo| 0.95  | 0.32  |
| sistema | --    | 0.49  |
| impacto | 0.18  | 0.38  |
| actividad| 1.14 | 1     |
| estrategia| 0.39 | 0.2   |
| necesidad| 0.37  | 0.22  |
| iniciativa| --  | 0.38  |

The second aspect that was taken into consideration is the proportion of “atypical” collocations in original and translated texts. In order to establish whether a collocation is “typical” or not, the list of the first 100 verbal collocates of each noun was observed in a reference corpus (for this task the Spanish corpus of the Leeds Collection of Internet Corpora was used, henceforth REF; Sharoff, 2006). This list was then compared with the nouns’ verbal collocates in OR-ES and TR-ES: if a verbal collocate used in one of the two corpora is also present in the list of the first 100 verbal collocates of the noun in REF, then that collocate is considered “typical”; conversely, a verb used in OR-ES and TR-ES that does not appear in the list of the first 100 verbal collocates of the noun in REF is considered “atypical”. Table 2 shows “atypical” verbal collocates for each corpus, with an indication as to whether they are used literally or metaphorically in each corpus (metaphorical collocates are in bold). Original and translated texts display exactly the same proportion of “typical” and “atypical” collocations, with the latter accounting for 14% of total collocations in both corpora (Figure 1). Atypical collocations in translated texts show a greater proportion of metaphorical uses than original texts (62% metaphorical uses in TR-ES vs. 50% in OR-ES).
Table 2. “Atypical” verbal collocates in each corpus.

|        | OR-ES | TR-ES |
|--------|-------|-------|
| emisión| --    | minimizar |
|        |       | estabilizar |
| programa| --    | ampliar |
|        |       | poner (en marcha) |
| riesgo | --    | abordar |
| proyecto| poner (en marcha) | poner (en marcha) |
| objetivo| integrar | -- |
| sistema | --    | estudiar |
| impacto| identificar | -- |
| actividad| --    | explorar |
| necesidad| dar (respuesta) | -- |

Collocation “typicality” was also investigated by comparing the ranking of collocates in OR-ES and TR-ES and the reference corpus. The focus was on the verbal collocates that show a significantly higher ranking in one of the two subcorpora at study when compared to REF. As shown in Table 3, compared with original texts, translated texts use a higher proportion of collocates with lower ranking in the reference corpus (11% of OR-ES collocates vs. 16% of TR-ES collocates have a higher ranking than in REF). This seems to suggest that translated texts tend to use more rare/“atypical” uses than original texts, a result that challenges the simplification and normalisation/conventionalisation hypotheses.

Table 3. Verbal collocates with a higher ranking in OR-ES or TR-ES than in REF.

|        | OR-ES | TR-ES |
|--------|-------|-------|
| negocio|       | desarrollar |
| emisión| compensar | generar |
|        | controlar |     |
| programa|       | apoyar |
|        |         | llevar |
|        |         | lanzar |
| riesgo |       | gestionar |
| desarrollo| generar |     |
| objetivo|       | superar |
| actividad|       | mostrar |
| iniciativa|       | poner (en marcha) |
3.2. Metaphorical conventionality

In order to compare the degree of metaphorical conventionality in original and translated texts, the first step was metaphor identification, which was carried out following the ‘Metaphor Identification Procedure (MIP)’ proposed by the (Pragglejaz Group, 2007). According to this procedure, a lexical unit can be considered metaphorical if (and where) it is used with a meaning that contrasts with an existing and identifiable more “basic” contemporary meaning of the word. Based on Conceptual Metaphor Theory tenets, a “basic” meaning is defined as: “more concrete (what they evoke is easier to imagine, see, hear, feel, smell, and taste); related to bodily action; more precise (as opposed to vague); historically older” (Pragglejaz Group, 2007: 3). An example is provided by the verb lanzar, which has a basic concrete, original, more precise meaning of “Referido a un objeto, darle impulso para soltarlo después, de modo que salga despedido con fuerza en una dirección” (CLAVE) (e.g. lanzar una piedra); however, the verb can also be used with a more abstract, less precise meaning which is not related to bodily action, indicating “hacer propaganda [esp. a una novedad] con una gran campaña publicitaria” (CLAVE) (e.g. lanzar un programa). This second, non-basic meaning is identified as metaphorical.

Based on these criteria, verbal collocates in OR-ES and TR-ES were classified as either literal or metaphorical. Findings show that translated texts tend to use a higher proportion of metaphorical uses (57%) than original texts, where metaphorical uses only accounted for 50% of verbal collocates. Moreover, translations display a higher proportion of repeated metaphorical verbs, with 37% of metaphorical verbs being repeated more than once. In original texts, repeated metaphors only account for 21% of total metaphorical expressions.

While these results seem to point to a more idiomatic, patterned use of language in translated texts, the proportion of metaphorical verbs that are shared by the two corpora points in the opposite direction. Original texts share 79% of their metaphors with translated texts, while for the latter the proportion of metaphorical verbs shared with original texts is only 44%. This means that 56% of all metaphorical uses in translated texts are specific to this corpus, and not found in original texts. Based on these findings, it would appear that there is a set of metaphors that belong exclusively to translated texts – which can be interpreted as an indication of the idiosyncratic nature of translations rather than a tendency to conventionalisation/standardisation.

A further aspect taken into consideration was metaphorical conventionality. This was defined using the criterion proposed by Deignan (2005) to distinguish between innovative and historical metaphors: a low frequency of metaphorical uses for a given word is considered indicative of high innovativity; conversely, if a word is almost exclusively used metaphorically and its literal uses are extremely rare, their metaphors can be considered as highly conventional.

To assess whether a metaphorical use is conventional or innovative, the frequency of metaphorical and literal uses for each verbal collocate was checked in the reference corpus. An example is provided by the verb phrase poner en marcha, for which two main senses were identified: the literal, concrete sense is used with direct objects indicating vehicle/machinery (“poner en marcha el coche donante”, REF), whereas the metaphorical sense is used with direct objects indicating initiatives/actions (“pone en marcha la búsqueda de Rosebud”, REF). To establish the
degree of metaphorical conventionality of the verb phrase, 100 randomly-selected concordance lines were checked in REF, finding that only 1 out of 100 uses of *poner en marcha* was literal. It was thus concluded that the verb phrase is used almost exclusively metaphorically and, therefore, is a very conventional metaphor.

This procedure was repeated for all the metaphorical verbs that appear in only one of the corpora (but not in the other). Table 4 contains the full list of metaphorical verbs taken in consideration for each corpus, with an indication of the number of literal uses they display in the reference corpus and a measure of the collocativity (as indicated by the log-likelihood measure) of the VERB-NOUN pair they appear in. A lower number of literal uses indicates greater metaphorical conventionality, while the log-likelihood measure is added to check whether non-conventional metaphorical verbs are used in very frequent collocations.

Data sparseness poses a serious problem for corpus comparison. The corpus of original texts only displayed 3 metaphorical verbs that are not used in translated texts. Of these, two are highly conventional (*impulsar* and *recoger*, with 5 and, respectively, 13 literal instances in REF). *Responder* shows a lower degree of conventionality (with 44 literal uses in REF). However, in the corpus it is used in the sequence *responder a las necesidades*, an extremely frequent collocation.

Metaphors in translated texts are all very conventional, with the less conventional ones appearing in very common collocations (e.g., *cubrir las necesidades*). The corpus, however, does contain a few non-conventional collocations, which are not found or are extremely rare in the reference corpus. These include: *abordar el riesgo; apoyar una estrategia; superar un objetivo; explorar las actividades*. In all these cases the metaphors used are very conventional; what makes the use less conventional is the specific VERB-NOUN collocation used.

Table 4. Metaphorical conventionality of the VERB-NOUN pairs.

| OR-ES | literal in REF | LL    |
|-------|----------------|-------|
| impulsar | 5              | 338.51 desarrollo |
| recoger | 13             | N/A   estrategia  |
| responder | 44             | 1002.77 necesidad |

| TR-ES | literal in REF | LL    |
|-------|----------------|-------|
| abordar | 0              | N/A   riesgo   |
| adoptar | 1              | 167.03 estrategia |
| apoyar | 1              | 54.42 programa |
| apoyar | 1              | 177.41 proyecto |
| apoyar | 1              | 121.65 desarrollo |
| apoyar | 1              | 23.97 estrategia |
| apoyar | 1              | 312.14 inicivativa |
| superar | 2              | N/A   objetivo |
| ampliar | 8              | 18.15 programa |
| explorar | 10             | N/A   actividad |
| introducir | 11           | 103.37 sistema |
| lanzar | 14             | 47.25 programa |
| lanzar | 14             | 55.16 proyecto |
| elaborar | 16             | 400.71 informe |
| elaborar | 16             | 239.06 programa |
| acelerar | 21             | 147.94 desarrollo |
| crear | 26             | 309.33 sistema |
| cubrir | 31             | 1299.13 necesidad |
4. Discussion

The aim of this paper was to explore and test a methodology to study the language of translations through a comparison of metaphor use in original and translated texts. One of the questions posed was whether metaphor use could confirm the observed tendency of translations to use a simplified, more explicit, and more conventionalised language than original texts.

This exploratory study compared metaphorical VERB-NOUN collocations in a monolingual comparable corpus of original and translated sustainability reports published by corporations operating in the banking and energy sectors. VERB-NOUN collocations were analyzed in order to compare: (a) variety and proportion of typical vs. atypical collocations; (b) degree of conventionality of metaphorical instances in original and translated texts.

Results do not straightforwardly confirm or contradict the Translation Universal hypotheses. On the one hand, VERB-NOUN collocations in translated texts yield a higher proportion of metaphorical instances with a marked tendency to repeat the same metaphors more often: this – in line with the normalization hypothesis – characterizes the language of translations as more idiomatic, standardized, and repetitive. On the other hand, however, translations also display greater collocational variety, more frequent use of collocations which are infrequent in native language, and a higher proportion of metaphorical instances which are specific to translated texts (and not found in original texts). All these aspects emphasize the idiosyncrasy of translated language, which – far from conforming to and amplifying typical patterns of language – shows a preference for atypical uses.

A possible explanation for this is provided by the notion of “shining through” (Teich, 2003), whereby source language features “shine through” or reverberate in the target text. This is definitely a factor at play in our corpus. An example is provided by the collocation abordar el riesgo found in our translated texts: while relying on a highly conventional metaphor (abordar), the expression creates an unusual collocation (Table 4 above) which results from the translation of the fairly frequent English collocation “address the risk” (59.89 log-likelihood in the English reference corpus). After checking the English original, it was found that 19 out of 27 instances of the English verb address were translated as abordar, while 5 of the remaining 8 uses of address were omitted in the Spanish translation. The atypical Spanish collocation may thus result from a (partially automated) “habit” to translate the verb address as abordar regardless of the nouns that collocate with it, thus creating a “shine-through” effect in the target text.

On the whole, the provisional data here illustrated suggest a combined influence of normalization and “shining through” in translated texts. The analysis has also provided methodological indications for future work: data sparseness has not allowed a quantitative comparison of metaphorical conventionality in original and translated texts; a problem that only a bigger corpus, and access to more data, can help solve. Furthermore, the criteria set for collocation selection proved too restrictive, as they are effective only in syntactically simple sentences/structures: future work will need to address this issue by expanding the word span used for collocation extraction. Finally, the study has shown that conventional metaphors can be expressed through “atypical” collocations, showing that there is not a direct link between metaphor conventionality and conventional language; in the future, a method should be devised that accounts for both metaphor and collocation conventionality.

References

Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), Text and Technology. In Honour of John Sinclair (pp. 233-250). Amsterdam/Philadelphia: John Benjamins.
Baker, M. (1996). Corpus-based Translation Studies: The Challenges that Lie Ahead. In H. Somers (Ed.), Terminology, LSP and Translation: Studies in Language Engineering: In Honour of Juan C. Sager (pp. 175-186). Amsterdam/Philadelphia: John Benjamins.
Bernardini, S. (2007). Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora. Proceedings of the Corpus Linguistics Conference (CL2007) (pp. 1-16). Birmingham: University of Birmingham.
Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. Proceedings of the 3rd Conference on Computational Lexicography and Text Research (pp. 23-32), Budapest, Hungary, 7th-10th July 1994.
Dayrell, C. (2007). A quantitative approach to compare collocational patterns in translated and non-translated texts. International Journal of Corpus Linguistics 12(3), 375-414.
Deignan, A. (2005). Metaphor and Corpus Linguistics. Amsterdam/Philadelphia: Benjamins.
Dunning, T. (1994). Accurate Methods for the Statistics of Surprise and Coincidence. Computational Linguistics, 19(1), 61-74.
Frawley, W. (1984). Prolegomenon to a Theory of Translation. In W. Frawley (Ed.), Translation: Literary, Linguistic and Philosophical Perspectives (pp. 159-175). London: Associated University Press.

Hansen-Schirra, S. (2011). Between normalization and shining-through. Specific properties of English-German translations and their influence on the target language. In S. Kranich, V. Becher, S. Höder & J. House (Eds.), Multilingual Discourse Production: Diachronic and Synchronic Perspectives (pp. 133–162). Amsterdam/Philadelphia: John Benjamins.

Kenny, D. (2001). Lexis and Creativity in Translation. A Corpus-Based Study. Manchester: St. Jerome Publishing.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43(4), 557-570.

Olohan, M. & Baker, M. (2000). Reporting ‘that’ in Translated English: Evidence for Subconscious Processes of Explicitation? Across Languages and Cultures 1(2), 141-158.

Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor & Symbol* 22(1), 1–39.

Samaniego Fernández, E., Velasco Sacristán, M. & Fuertes Olivera, P. A. (2005). Translations we live by: The impact of metaphor translation on target systems. In P. A. Fuertes Olivera (Ed.), Lengua y Sociedad: Investigaciones recientes en Lingüística Aplicada (61-81). Valladolid: University of Valladolid.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In D. Jones (Ed.), International Conference on New Methods in Language Processing (pp. 44-49), Manchester, UK, September 1994, Manchester, University of Manchester Institute of Science and Technology (UMIST).

Teich, E. (2003). Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts. Berlin: Mouton de Gruyter.

Sharoff, S. (2006). Creating general-purpose corpora using automated search engine queries. In M. Baroni & S. Bernardini (Eds.), WaCky! Working papers on the Web as Corpus (pp. 63-98). Bologna: Gedit.