

The idiom principle and translated texts

Gill Philip (University of Macerata, Italy)

Introduction

The idiom principle has been a mainstay of (monolingual) phraseological research for the past quarter century. First advanced by Sinclair (1991), it proposes that normal text is composed of intersecting pre-fabricated chunks, only occasionally necessitating a switch to allow for the insertion of terminological (*ibid.*) insertions before switching swiftly back again to the phrasal norm. Observable in corpus data via KWIC concordances – where minor variations can be absorbed – as well as through n-gram and cluster extraction which have more stringent requirements regarding fixity of word-form and syntactic sequencing, it is clear that the idiom principle does indeed hold sway in normal language use. But what happens when a text is not being constructed, *per se*, but rather *reconstructed*, as is the case with translated texts? Little is known about the extent to which the idiom principle survives the process of translation because until now very few scholars have investigated multiple translations, and none of those who have done so have investigated phraseological (rather than terminological) phenomena.

This contribution seeks to redress this gap in knowledge, drawing on the author's own learner translation corpus. Compiled for use in the classroom, the data consists of the students' translation assignments: in other words, multiple translations of selected source texts. These are subjected to linguistic analysis via concordancing software in the classroom, allowing students to view all versions – their own and those of their classmates – via projected concordances called up in real time. Concordancing multiple translations makes it possible to identify points of convergence (identical or similar translation solutions) and divergence (wide variety of translation solutions). Convergence reveals that the class has reached a tacit consensus on how best to translate an item or text fragment, with no evident problems arising. Divergence, in contrast, is indicative of a translation problem (such as misunderstanding of the source text or problems in rendering the meaning effectively in the target language). Unsurprisingly, it is on translation problems that most classroom discussion hinges; but beyond the immediate needs of the translation class, convergence too deserves close linguistic scrutiny. The dearth of research on multiple translations prevents us from assessing at present just how similar translated texts ought to be – they are, after all, reconstructions of an existing text in another language. What we can investigate, instead, are the ways in which convergence manifests itself, and how these relate to existing theoretical models of 'normal' text production.

Data and methods

The corpus used in this study consists of ten small subcorpora, each comprising between ten and fourteen translations into Italian (L1) of a 400-word English (L2) source text. Since at present no parallel concordancer can handle so many texts, AntConc (Anthony 2014) sufficed for the extraction of items to study in class. With no text alignment available, the search procedure relies on unambiguous items in the source text, e.g. proper names, dates and figures, etc., which serve as entry points into the data. Using the widest context span available, it is then possible to locate the item of interest on-screen.

The idiom principle, convergence, and divergence in multiple translations

When analysing the translations, it was found that convergence is a strongly phraseological phenomenon: not only were the translations of significant content words agreed upon, but so too was their phraseological environment. While this might be expected to some extent, given that we are dealing with translation rather than original text composition, this feature is of particular interest to linguists and translators alike. By way of background, it is worth noting that while the morphological simplicity of English lends itself well to strong phraseological patternings, highly inflected languages such as Italian are expected to perform less well: each verb has six distinct forms for each of its eight tenses, all nouns inflect for gender and number, and adjectives agree with these, having up to four possible forms each (masculine singular/plural; feminine singular/plural); the definite article has six forms including the contraction *l'*, and prepositions can be merged with these as *preposizioni articolate*. Not requiring the specification of subject pronouns, word order too is often flexible: the object is often placed before rather than after intransitive verbs, and in translation, pronouns can be specified (mimicking the source text) or omitted, at will. Morphological variety and syntactic flexibility therefore reduce the likelihood of recurring patterns, particularly n-grams, being found – even in multiple translations of a single source text.

Considering this potential phraseological variability, we may well expect to find 3- and 4-grams in abundance, but few longer ones. However, the cases of phraseological convergence found in the data were not limited to shorter n-grams. Counting only those found in at least half of the translations for each source text (i.e. Range= min. 5), n-grams of five to eight consecutive words were common, and longer strings were often found. For example, in one of the texts, a 400-word extract from Packard (1957), there were twenty-seven 8-grams, seventeen 9-grams, ten 10-grams, five 11-grams, two 12-grams, and one 13-gram. Only one of the 8-grams occurred in all ten translations (compared to twenty-five 3-grams, thirteen 4-grams, eight 5-grams, five 6-grams, and three 7-grams), but this still leaves the fact that between five and nine independently-worked translations had many long clusters in common, providing evidence of common phraseological usage in these students' use of their native tongue. The causes of this are yet to be unravelled: source text influence is likely to be a factor, but not the only one.

Extending the notion of phraseological regularity to include minor variation, convergence in translation becomes even more marked. It emerges clearly when viewing the translations on-screen via a concordance, and its pervasiveness confirms that these student translators largely agree on how to translate their texts. There may be minor differences in word choice, or verb tense, differences which in turn impose morphological or grammatical limitations on the immediate textual environment, but this is visibly different from divergent translations which conform to the features associated with the open choice principle (Sinclair, 1991): there is a sudden switch from the phraseological flow of convergent translations to divergence; and the switch back again is just as sudden.

Figure 1 shows an example from Packard (1957), which demonstrates first of all moderate convergence (a variety of forms of the verbs *usare* and *utilizzare*, to translate *use*), followed by clear divergence (the translation of *tossed salad*), returning again to convergence at the end of the fragment. *Tossed salad* – not in itself a difficult term – is evidently difficult to translate in this text: is not translated at all in translations 7-10, indicating that students were unable to interpret its meaning in context. Translations 1 and 2 simply maintain the original. Elsewhere we can see indicators of uncertainty: queries in translations 3 and 4, scare quotes in translation 7. Translations 4, 5, and 6 modify the original item by explicitation (Baker, 1993) and the student responsible for

translation 6 further explicitates it (erroneously) as being metaphorical. Yet despite the evident difficulty that this item posed, the phraseology returns to normal immediately afterwards, embodying the dramatic switch between open choice and idiom principles that Sinclair (1991: 110) hypothesized.

1. ...utilizzò		l'insalata	come indicatore più affidabile...
2. ...ha usato		l'insalata	come un più affidabile indicatore...
3. ...ha utilizzato		l'insalata (???)	come un indicatore più affidabile...
4. ...utilizzò		l'insalata mista (?)	come indicatore ben più attendibile
5. ...utilizzò		l'insalata già pronta	come un indicatore più affidabile...
6. ...utilizzò		la metafora dell'insalatona	come indicatore più affidabile...
7. ...ha usato	la	"tossed salad"	come indicatore più affidabile...
8. ...usò		[<i>tossed salad</i>]	come un indicatore più affidabile...
9. ...usa	[<i>the</i>	<i>tossed salad</i>	<i>as a more reliable indicator...</i>
10. ...used	<i>the</i>	<i>tossed salad</i>	<i>as a more reliable indicator...</i>

Figure 1. Translations of "...used the tossed salad as a more reliable indicator of..."

Concluding remarks

This contribution will further examine cases of idiom and open choice principle in the author's multiple variant learner translation corpus, drawing attention to the role of the idiom principle in translation, not only as product, but also as a text (re-)creating process.

References

- Anthony, L. (2014). *AntConc* (Version 4.3.4). Tokyo: Waseda University.
- 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: John Benjamins.
- Packard, V. (1957). *The Hidden Persuaders*. London: Penguin.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.