SOME PROBLEMS OF TRANSLATING IDIOMS

Ümit Öz

Catedra Filologie Engleză

În articolul dat, sunt analizate problemele ce țin de traducerea unităților frazeologice. Sunt evidențiate dificultățile lexical-stilistice în traducerea frazeologismelor.

An idiom is an expression (i.e. term or phrase) whose meaning cannot be deduced from the literal definitions and the arrangement of its parts, but refers instead to a figurative meaning that is known only through conventional use. In linguistics, idioms are widely assumed to be figures of speech that contradict the principle of compositionality, however some debate has recently arisen on this subject.

In the English expression *to kick the bucket*, a listener knowing only the meaning of kick and bucket would be unable to deduce the expression's actual meaning, which is to die. Although kick the bucket can refer literally to the act of striking a bucket with a foot, native speakers rarely use it that way.

Idioms hence tend to confuse those not already familiar with them; students of a new language must learn its idiomatic expressions the way they learn its other vocabulary. In fact many natural language words have idiomatic origins, but have been sufficiently assimilated so that their figurative senses have been lost.

Idioms are, in essence, often colloquial metaphors – terms which require some foundational knowledge, information, or experience, to use only within a culture where parties must have common reference. As cultures are typically localized, idioms are more often not useful for communication outside of that local context. However some idioms can be more universally used than others, and they can be easily translated, or their metaphorical meaning can be more easily deduced.

The most common idioms can have deep roots, traceable across many languages. To have blood on one's hands is a familiar example, whose meaning is obvious. Many have translations in other languages, some of which are direct. For example, get lost! (ie. go away or stop bothering me) is said to have originated from a Persian expression., "gom sho!" which means, quite literally, "become lost."

While many idioms are clearly based in conceptual metaphors such as "time as a substance", "time as a path", "love as war" or "up is more", the idioms themselves are often not particularly essential, even when the metaphors themselves are. For example "spend time", "battle of the sexes", and "back in the day" are idiomatic and based in essential metaphors, but one can communicate perfectly well with or without them.

In forms like "profits are up", the metaphor is carried by "up" itself. The phrase "profits are up" is not itself an idiom. Practically anything measurable can be used in place of "profits": "crime is up", "satisfaction is up", "complaints are up" etc. Truly essential idioms generally involve prepositions, for example "out of" or "turn into".

Interestingly, many Chinese characters are likewise idiomatic constructs, as their meanings are more often not traceable to a literal (i.e. pictographic) meaning of their assembled parts, or radicals. Because all characters are composed from a relatively small base of about 214 radicals, their assembled meanings follow several different modes of interpretation - from the pictographic to the metaphorical to those whose original meaning has been lost in history.

In this article we will analyse the treatment of fixed word expressions developed the translation system. We will show several cases of transfer to corresponding idioms in the target language, or to simple lexemes in Romanian and English.

Translating idioms is one of the most difficult tasks for human translators and translation machines alike. Idioms are defined as multiword expressions with a fixed (often metaphorical) meaning that cannot be derived from its parts. It is one of the most frequently used means of non-literal language.

Literal translation of idioms is a source of numerous translator's jokes and apocrypha. The following famous example has often been told both in the context of newbie translators and that of machine translation: when the sentence "The spirit is strong, but the flesh is weak". Was translated into Russian and then back to English, the result was "The vodka is good, but the meat is rotten".

Idioms can be classified in various ways. They can, for example, be distinguished by their syntactic structure as in 1. These examples show that some idioms can be translated word by word if a similar idiom in the target language exists (the verb phrase example), while others can be translated using the same picture but with a different structure (the infinitival complement example), and still others cannot be translated with an idiom but only with their literal meaning if a corresponding idiom does not exist in the target language (the noun phrase example).

1) noun phrase: *a broad hint*
verb phrase: *throw the baby out with the bathwater*
infinitival complement: *without batting an eyelid*

Idioms can also be distinguished by their degree of compositionality. We distinguishe three classes of idioms:

2) compositional: *have a good (bad) hand*
partly compositional: *to watch something like a hawk*
non-compositional: *not to do things by halves*

A compositional idiom has two characteristics: First, it can be syntactically modified and second its parts can be mapped to the intended meaning. In a partly compositional idiom at least one constituent has its original meaning whereas the rest has a special idiomatic meaning. In example 2 *to watch* has its genuine meaning whereas *with the eyes of Argus* is specific to this idiom. The noun *Argus* is not used outside of this idiom. It is a further characteristic of idioms that they use specially preserved lexical material. A non-compositional idiom can be neither syntactically modified nor lexically substituted without losing its idiomatic meaning.

A translation system must recognize idioms and translate them as a whole. This should be easiest for non-compositional and partly compositional idioms since they are fixed in their lexical material. It is more difficult for compositional idioms since their variations must be taken into account.

Idioms can be contrasted to *collocations*. Collocations are also relatively fixed combinations of words but their meaning can be derived from their parts. It is the special combination of words and their frequent co occurrence rather than their special meaning that sets collocations apart from idioms.

Multiword expressions are known to constitute a serious problem for natural language processing. In the case of translation, a proper treatment of multiword expressions is a fundamental requirement, as few customers would tolerate a literal translation of such common expressions as *a intra în vigoare* 'to come into effect', *a da dovadă* 'to show' or *a face cunoştinţă* 'to meet '.

However, a simple glance at some of the current commercial translation systems shows that none of them can be said to handle multiword expressions in an appropriate fashion. As a matter of fact, some of them explicitly warn their users not to use multiword expressions.

We distinguish verbal idioms (idioms headed by a verb) from fixed multiword entries: the latter, in our definition, cannot be discontinuous, and are stored like ordinary words in our monolingual dictionary. Examples include nouns such as *a înnoi* (‘update’) and *a-şi asuma răspunderea* (‘support’ in our technical corpus), conjunctions such as *de manieră că* (‘so that’), and prepositions (*la sfârşit de* ‘at the end of’). These expressions do not require a specific treatment, and are handled in the same way as single words of the same category.

For the purpose of this paper, we define verbal idioms as verb phrases whose meaning is idiomatic and cannot be derived compositionally from the literal meaning of the idiom parts. Verbal idioms thus pose problems for natural language systems, and especially machine translation systems, where the entire phrase may have a non-compositional gloss. For example, in the system presented in this paper, the Romanian idiom *a-şi asuma răspunderea* has been variously translated word for word (and therefore incorrectly) as ‘take in load’ or ‘seize in load’, when the correct translation in our technical context is ‘support’.

Other examples include *a face parte din* (‘belong’), translated by ‘make part’ in some instances, *a avea nevoie de* (‘need’) translated by ‘have need’, etc. These verbal idioms are very common, and are typically translated very poorly. The problem in all these cases is that these verbal idioms are not analyzed as such, and are translated literally, word for word.

Verbal idioms can participate in a variety of constructions, which can result in discontinuities, and vary according to the idioms [1,2,3 among others]. That, in turn, makes it difficult to match all parts of the idiom in a sentence. For example, in the examples below, the object is not adjacent to the other idiom parts because of relativization (as in (1)) and the passive construction (in (2)):

(1) *Afacerile de care şi-a asumat răspunderea sunt delicate.*

(2) *Aceste afaceri vor vi luate sub răspundere fără întârziere.*

The entire expression has to be recognized as a unit, however, if translations such as the ones mentioned in the introduction are to be avoided. For the expression to have an idiomatic reading, *a lua* must be followed by, although not necessarily be adjacent to, the prepositional phrase *sub răspundere*; it must also have an object complement (which, in the passive construction, will be realized as the grammatical subject).

Previous approaches to idiom analysis propose to identify idioms during parsing (for example [4,5]), or on the structure produced by parsing [6]. Some approaches propose local grammar rules written specifically to handle idioms [7].

Our approach is closest to Wehrli's solution, in that idioms are identified after parsing (in our case, on the resulting syntactic tree). As we pointed out earlier, since idioms can be discontinuous, the entire sentence has to be parsed before an idiom can be identified with certainty. In our current research, idioms such as these are entered manually in the monolingual dictionaries. The entries are keyed on the verbal head, and they list the arguments and modifiers that make up the idiom, with morphosyntactic constraints expressed as features on each idiom part. For example*, a avea nevoie de (ceva)* is an idiom meaning 'to need (something)' as long as *nevoie* is in the singular and is not preceded by a determiner; that information is hand-coded in the dictionary.

The translation of these idioms is a three-step process:

- Identification of source idiom
- Transfer of idiom
- Generation of target idiom

Idiom identification

As we argued in the previous section, the task of identifying an idiom is best accomplished at the abstract level of representation. At this point, the structure is completely general, and does not contain any specification of idioms. The idiom recognition procedure is triggered by the "head of idiom" lexical feature. This feature is associated with all lexical items which are heads of idioms in the lexical database.

The task of the recognition procedure is to retrieve the proper idiom, if any, and to verify that all the constraints associated with that idiom are satisfied.

Idiom entries specify the canonical form of the idiom (mostly for reference purposes), the syntactic frame with an ordered list of constituents, and the list of constraints associated with each of the constituents.

Transfer and generation of idioms

Once properly identified, an idiom will be transferred as any other abstract lexical unit. In other words, an entry in our bilingual lexicon has exactly the same form no matter whether the correspondence concerns simple lexemes or idioms. The corresponding target language lexeme might be a simple or a complex abstract lexical unit. For instance, our bilingual lexical database contains, among many others, the following correspondences:

Romanian	**English**
a avea nevoie de X	*need X*
a face cunoştinţă cu X	*meet X*
a avea dorinţa de	*feel like*
ce muscă a pişcat	*what has gotten*

The generation of target language idioms follows essentially the same pattern as the generation of simple lexemes. The general pattern of generation is the following: first, a maximal projection structure is projected on the basis of a lexical head and of the lexical specification associated with it. Second, syntactic operations apply on the resulting structure (extraposition, passive, etc.) triggered either by lexical properties or general features transferred from the source sentence. For instance, the lexical feature [-(-raising] associated with a predicate would trigger a raising transformation (NP movement from the embedded subject position to the relevant subject position). Subject-Auxiliary inversion, topicalization, auxiliary verb insertion are all examples of syntactic transformations triggered by general features, derived from the source sentence.

The first step of the generation process produces a target language D-strcuture, while the second step derives S-structure representations. Finally, a morphological component will determine the precise orthographical/ phonological form of each lexical head.

In the case of target language idioms, the general pattern applies with few modifications. Step 1 (projection of D-structure) is based on the lexical representation of the idiom (which specifies the complete syntactic pattern of the idiom, as we have pointed out earlier), and produces structure. Step 2, which only concerns the insertion of perfective auxiliary, derives the S-structure. Finally, the morphological component derives sentence.

In this section, we have argued for a distinct treatment of compounds, viewed as complex lexical units of word-level category, and of idioms, which are phrasal constructs. While compounds can be easily processed during the lexical analysis, idiomatic expressions are best handled at a more abstract level of representation, in our case, the D-structure level produced by the parser. The task of recognition must be based on a detailed formal description of each idiom, a lengthy, sometimes tedious but unavoidable task. We have then shown that, once properly identified, idioms can be transferred like any other abstract lexical unit. Finally, given the fully-specified lexical description of idioms, generation of idiomatic expressions can be achieved without ad hoc machinery.

References:

1. Nunberg G. Idioms. Language 70:3:491-538. I. Sag & T. Wasow, 1994.
2. Schenk A. The syntactic behaviour of idioms // Everaert M., van der Linden E., Schenk A. and Schreuder R. (eds). Idioms: Structural and Psychological Perspectives, 1995.
3. Wehrli E. Translating idioms // Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics. - Montreal, 1998.
4. Stock O. Parsing with flexibility, dynamic strategies, and idioms in mind // Computational Linguistics. - 1989. - 15.1.
5. Matsumoto Y. et al. Bi-directional parsing for idiom handling // Martin J., Fass D. and E. Hinkelman (eds). Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language: Metaphor, Metonymy, Idioms, Speech Acts and Implicature, 1991.
6. Wehrli E. Op. cit.
7. Breidt E. et al. Local grammars for the description of multi-word lexemes and their automatic recognition in texts. Proceedings of COMPLEX 96. - Budapest, 1996.

Prezentat la 24.04.2007