Multilingual idioms and proverbs: Wiki as a tool in collaborative translation

Dora Montagna
Unywersytet Warszawski - Katedra Italianistyki
ul. Oboźna 8
00-332 Warszawa
+48 225520697
dora.montagna@gmail.com

ABSTRACT
This work aims to provide a useful tool to translators, linguists and students who want to study and find equivalences of idioms and proverbs in different languages. The goal is to create an extensive multi-word expressions and proverbs corpus, overcoming the main problem of their classification, thanks to hypertext and key-word based search: a synthesis of a glossary and a forum, easily consulted; open to each external contribution, in regard to the changeable nature of such linguistic heritage.

Categories and Subject Descriptors

General Terms
Languages

Keywords
Paremiology, phraseology, multi-word expressions.

1. INTRODUCTION
Idioms represent a window on a community’s linguistic culture. Proverbs, from which idioms often draw on, are expressions of popular wisdom, which is why they exist in a big variety of forms and nuances. It is evident that such expressions represent a great difficulty for a translator. But the cultural debate that emerges from their study involves translators, linguists and the whole community of speakers. Such oral knowledge, for its changeable nature, is hardly held and registered in collections and conventional dictionaries. For this reason the Wiki as a collaborative and open tool, seemed to me the ideal instrument to open the confrontation.

2. A WIKI FOR TRANSLATION
With this work I would like to present a method of computer-aided translation concerning a very specific aspect of the language: idioms and proverbs. Such method is based on open confrontation and collaborative work, provided by a Wiki-based application1. MWEs (multi-word expression) are traditionally collected and classified under different criteria, not always objective. This application is aimed to create an extensive multi-word expressions and proverbs corpus, overcoming the main problem of their classification, thanks to hypertext and key-word based search. Another crucial problem for the translator is to get the sense and the proper use of the MWE. It may concern not only idiomatic MWEs and proverbs, but also technical and special language MWEs. That’s why a cooperative tool, such as a forum, is essential. This tool is also useful in language teaching and learning, as an instrument of collaborative investigation on the cultural background of a community, and its folklore and traditions.

2.1 Corpus

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. WiKiSym ’08, September 8-10, Porto, Portugal
Copyright 2008 ACM 978-1-60558-128-3/08/09...$5.00.

Figure 1. Key-word index.

1 The corpus can be consulted at http://www.dga-due.info/index.php/Main_Page. Should the access be denied, because of work in progress, please contact with the author.
This application presents a multilingual extensive corpus of idioms and proverbs as a database. In order to limit the domain of investigation, I registered in my corpus all expressions containing the word “mosca / mucha” (“fly”, in English) in Italian, Spanish and Polish. I collected all entries registered in mono and multilingual dictionaries, considering historic collection as well, and I tried to establish equivalences under specific criteria². Thanks to the elastic structure of Wiki, the corpus can be organized in a multiple way without interfering with its consultation.

2.2 Search for Entries

This application contains a multiple index (one index for each language) which links in a hypertextual way to registered entries, and also allows a key-word based search. Moreover, each entry of the corpus is codified, as well as each page of the Wiki, in order to always be increased maintaining the same classification criterion. When selecting a MWE or proverb, a new page containing all variants and matching expressions in other languages is displayed. Users can contribute to this page by adding further variants (e.g. local variants) or equivalents expressions in other languages, since they change depending on the geographical, historical and social position.

Figure 2. Multilingual MWEequivalents.

2.3 Discussion Page

Moreover, it is possible to enter all comments concerning sources and linguistic and translating issues in another page linked to each entry, useful to the translator in order to contextualize the expression and to select the most adequate variant. Such information is very difficult to obtain by traditional dictionaries and without the help of a mother tongue speaker. In this way, each linguistic and cultural commentary, observation and annotation is displayed. It is important to maintain an open access to modification of idioms and proverbs corpora so each user can contribute to their interpretation and add new expressions and new variants. A collaborative and dynamic environment, such as the Wiki, reflects well the mutable characteristic of the language concerning idiomatic and proverbial expressions.

Further developments of this work are taking into consideration different schemes for a corpus classification and annotation, aimed to make the structure of the Wiki accessible to machines as well. In collaboration with a research team from ENEA (Italian National Agency for New Technology, Energy and Environment), we are creating a system of concept trees based on top concept ontology such as MultiWordnet. We are also going to study FrameNet based relations and other systems, aiming to consider the most appropriate concept model for MWE and metaphors classification.

3. ACKNOWLEDGMENTS

I would like to thank Dr. Luciana Bordoni, from ENEA (Advanced Physical Technologies and New Material Department-Rome) for her helpful suggestions and her support.

4. REFERENCES


² The study on proverbs was included in the paremiology project “Gnome” (http://www.gr-gnome.eu/).
³ A research group of the Italian Filology Department (Katedra Italianistyki) of the UW is conducting a survey in order to establish a grade of effective use and knowledge for each MWE.