Elisabeth Piirainen (Steinfurt, Germany)

MWEs from a multi-lingual perspective: New insights into syntactic, lexical and semantic parallels of multiword expressions in diverse languages

The topic of this chapter is multiword expressions that occur in a large set of languages in almost the same morpho-syntactic and lexical structure and non-literal core meaning. The starting point is the project "Widespread Idioms in Europe and Beyond" which has access to multiword expressions (mainly idioms) from 78 European and 20 non-European languages. In contrast to traditional cross-linguistic phraseological studies, whose goal is to bring out the subtle differences between phrasemes of two or more languages, the focus of this project is to show the parallels of phrasemes in many languages. For the theoretical framework, the definition of "widespread idioms" (WIs for short), and the criteria to single them out see Piirainen (2012: 59–71). In this chapter, we restrict ourselves to the small group of weakly figurative widespread MWEs. The first section will outline the goals and methodological approaches of our study. The following section will be devoted to the explication of syntactic-lexical patterns which vary gradually in view of their far-reaching similarities, while the last section will summarize some results in the context of a more general MWE research.

1. Goals and methodological approaches

One goal of the widespread idiom project is to systematically single out those MWEs that have parallels in a large variety of geographically distant and genetically diverse languages from other MWEs (whose circulation may be restricted to a few languages of a small area). Completeness is sought: By means of the currently available options for networking it is possible to completely identify the relevant widespread MWEs. Reliable empirical research is required. About 350 participants of the project contributed to the data collection by filling in a number of questionnaires and reviewing the data with the help of text corpora of their native languages.

The next goal is the exact documentation of the identified common MWEs to make them available for further research. We are guided exclusively by the multi-language data themselves, without establishing an a priori theory. The cross-linguistic parallels but also possible gaps, deviations and differences will be checked carefully, including references to the text behavior of the MWEs. It cannot be excluded that the many near-equivalents may show – from a functional point of view – subtle pragmatic nuances and diasytematic differences. To uncover these dissimilarities still remains the usage-based research of individual philologies.

The most challenging question is that about the causes of the wide spread usage of MWEs across many languages, and it can now be partly answered, with all due caution. However, here we leave the realm of synchronic linguistics, because extralinguistic factors need to be taken into account. Often, intertextuality can be recognized as one of the reasons of the spread (for example, MWE (ii) below originates from a once well-known fable of Aesop). Rarely, polygenesis can be assumed, as for (iii) below which is based on general human experiences. However, we will not dwell on these questions here.

2. Parallel syntactic-lexical patterns in MWEs of diverse languages

Let us look at some examples. All of them come from the target domain TIME and are only weakly idiomatical. However, they show distinguished morpho-syntactic and lexical structures across a large variety of languages, a fact that enables us to design appropriate patterns (following the procedure of construction grammar). We can establish a typology of similarities, ranging from the complete uniformity in all languages to more or less competing groups of syntactic and lexical variants. We can only present a small part of our data.

(i)   *now or never* 'it must be done now or not at all; let us seize on the opportunity now'

MWE (i) is an example of equivalences across the board; there is only one pattern for all units reported by our respondents:

> [NOW – OR – NEVER]: Icelandic *nú eða aldrei*, Luxembourgish *elo oder ni*, Irish *anois nó choice*, Spanish *ahora o nunca*, Romanian *acum sau niciodată*, Latvian *tagad vai nekad*, Russian *сейчас или никогда*, Polish *teraz albo nigdy*, Croatian *sad ili nikad*, Bosnian *sad ili nikad*, Bulgarian *сега или никога*, Greek *τώρα ή ποτέ*; Armenian *ayzhm kam yerbek'*, Hungarian *most vagy soha*, Finnish *nyt tai ei koskaan*.

(ii) *slowly but surely* 'not fast, but inexorably; gradual but certain to finish'

MWE (ii) also shows a high degree of correspondences. Variants are restricted mainly to the use of the second adverb. The construction model [adv_1 – but – adv_2] manifests itself in two versions:

(ii)₁ **[SLOWLY but SURELY]**: German *langsam aber sicher*, Welsh *yn araf ond yn sicr*; Catalan *lenta però segur*; Lithuanian *lėtai, bet užtikrintai*; Serbian *полако али сигурно*, Albanian *ngadalë por sigurt*, Hungarian *lassan, de biztosan*, Finnish *hitaasti mutta varmasti*, Tatar *әкрен, ләкин дөрес*

(ii)₂ **[SLOWLY but STEADY]**: North Frisian *eewen, man aleewen*, Greek *αργά αλλά σταθερά*

(iii) *night and day* 'all the time; around the clock, perpetually'

MWE (iii) is extremely widespread. It has been found to exist in almost all of the European languages we analyzed and far beyond, even in Wolof, Yoruba or Māori. Variants extend to the constituent inversion, which is normal for similar binomials across languages, and to the omission of a copula. So there are – slightly simplifying – four patterns:

(iii)₁ **[NIGHT and DAY]**: Norwegian *natt og dag*, Portuguese *noite e dia*, Udmurt *уй но нунал*, Maltese *lejl u nhar*

(iii)₂ **[DAY and NIGHT]**: Irish *lá agus oíche*, French *jour et nuit*, Russian *день и ночь*, Estonian *päeval ja ööl*, Georgian *dgh'e da gh'ame*; Basque *gau eta egun*

(iii)₃ **[NIGHT – DAY]**: Breton *noz-deiz*, Finnish *yötä päivää*, Mari *jydshö-ketsyzhe* Turkish *gündüz gece*

(iii)₄ **[DAY – NIGHT]**: Aromanian *dzuuùã-noa*, Greek *μέρα νύχτα*, Moksha Mordvin *шинек-венек*, Tatar *көнен төнен*, Kazakh *күні-түні*

(iv) *one fine day* 'a particular day that one cannot determine; on an unspecified day, some time'

This MWE seems to be at best a weakly figurative collocation. However, it differs significantly from free word combinations, i.e. by the word *fine* which does not encounter in its true meaning but in quite a different one. Equivalents in various languages just follow this same lexical-syntactic structure but differ in view of the adjective:

(iv)₁ **[(on) – ONE – FINE – DAY]**: English *one fine day*, Hungarian *egy szép nap*

(iv)₂ **[(on) – ONE – BEAUTIFUL – DAY]**: Faroese *ein vakran dag*, Dutch *op een mooie dag*, Latvian *kādā jaukā dienā*, Slovak *jedného pekného dňa*, Greek *και μια ωραία μέρα*, Georgian *ert mshvenier dghes*

(iv)₃ **[(on) – ONE – GOOD – DAY]**: West Frisian *op in goede dei*, French *un beau jour*, Italian *un bel giorno*, Catalan *un bon dia*, Albanian *nje dite te mire*

(v) *from time to time* 'occasionally, now and then'

Most MWEs in (v) are consistent in form and lexis and follow one single pattern, but there are variations in the North Germanic and the Turkic languages:

(v)₁ **[FROM – TIME – TO – TIME]**: German *von Zeit zu Zeit*, Irish *ó am go ham*, Spanish *de tiempo en tiempo*, Latvian *laiku pa laikam*, Russian *время от времени*, Czech *čas od času*, Slovak *z času na čas*, Russian *время от времени*, Greek *από καιρό σε καιρό*, Hungarian *időről időre*, Karaim *vachttan vachtcha*

(v)₂ **[FROM – TIME – TO – THE OTHER]**: Norwegian *fra tid til annen*, Swedish *från tid till annan*, Danish *fra tid til anden*

(v)₃ **[TIME – TIME]**: Turkish *zaman zaman*, Tatar *вакыт-вакыт*, Bashkir *вакыт-вакыт*

3. Outlook

Similarities among a number of European languages have been well-known for a long time. They have been studied at different linguistic levels, be it phonetics/phonology, morphology or syntax, including similarities of genetically unrelated and geographically or culturally distant languages, compare the comprehensive EUROTYP research on linguistic typology whose results are based on morphologic and syntactic data from eventually 129 languages. The level of multiword expressions, however, has been excluded almost completely from all attempts of multi-lingual research. MWE research disciplines (such as phraseology) usually restricted themselves to comparing two or three languages and did not look at the linguistic situation of Europe as a whole. Recent research on "widespread idioms" has shown that the similarities in the realm of multiword expressions are much greater than previously thought. There has been little information so far on the occurrences of nearly identical MWEs across a large variety of languages. These findings are new and enable us to revise earlier views of the multiword lexicon of different languages. We do hope that the study of widespread MWEs could provide useful insight for empirical studies, such as phraseography, and, last but not least, contribute to a Theory of Multiword Expressions.

References:
Piirainen, Elisabeth. 2012. *Widespread Idioms in Europe and Beyond. Toward a Lexicon of Common Figurative Units*. New York [etc.]: Peter Lang.