

The Explanatory Combinatorial Dictionary as the Key Tool in Machine Translation

Igor Mel'čuk

Department of linguistics and translation
University of Montreal
C. P. 6128, Succ. Centre-Ville
Montreal (Quebec) H3C 3J7
CANADA
Igor.Melcuk@umontreal.ca

Abstract. Lexical and syntactic mismatches between languages are the main challenge for every translation (especially formidable for Machine Translation; see Mel'čuk & Wanner 2001, 2006). An example of mismatches in English-to-Russian translation:

a. *The demonstrators were brutally beaten and tear-gassed by the police.*

b. Rus. *Demonstranty podverglis' zverskomu izbieniju so storony policii; protiv nix byl primeněn slezotočivyj gaz* lit. 'Demonstrators underwent bestial beating from_side of_police; against them was applied tear-gas'.

The only way to resolve inter-linguistic mismatches is paraphrasing—intra- or interlinguistic. To explain better what I mean I will refer to mismatches found in the example. Thus, in Russian:

— The verb with the meaning '[to] tear-gas' does not exist, so you have to use the semantically equivalent expression '[to] apply tear-gas'.

— The passive of the verb '[to] beat'—with the Actor 'police'—does not readily combine with the noun 'police' as the syntactic Agent (it is not official enough!), so you have to use the semantically equivalent expression '[to] undergo a beating' (with the Subject 'demonstrators').

— The expression '[to] apply tear-gas' does not have a passive and therefore cannot be directly conjoined with '[to] undergo a beating' (the two clauses have different Subjects), so you have to use loose coordination of two complete clauses.

Such paraphrases are described by Deep-Syntactic Paraphrasing System of the Meaning-Text Theory: a few dozens universal paraphrasing rules (Mel'čuk 1992; Milićević 2006).

Paraphrasing at the semantic and/or deep-syntactic level can be ensured only by an extremely rich dictionary. Thus, the transformation '[to] tear-gas' \Rightarrow '[to] apply tear-gas' is triggered by the information in the lexical entry of the Russian nominal expression SLEZOTOČIVYJ GAZ 'tear-gas' (where the semi-auxiliary verb PRIMENJAT' '[to] apply' is found).

Such a dictionary, semantically-based and lexical co-occurrence centered, is the *Explanatory Combinatorial Dictionary* (Mel'čuk & Žolkovskij 1984, Mel'čuk *et al.*

2 Igor Mel'čuk

1984-1999, Mel'čuk *et al.* 1995). Three major zones of its lexical entry for headword L present:

— L's meaning, in the form of a formalized analytical lexicographic definition equivalent to a semantic network of the Meaning-Text Theory;

— L's Government Pattern, which contains complete information about L's actants and their surface realization;

— L's restricted lexical cooccurrence, in the form of Lexical Functions, which supply complete information about L's collocations (for instance: $Real_1(\textit{slezotočivj\ gaz}$ 'tear-gas') = *primenjat'* [\sim *protiv* 'against' N_V]; [$AntiBon + Magn$](*izbivat'* '[to] beat') = *zverski* lit. 'bestially'; $Fact_2(\textit{policija}$ 'police') = *izbivat'* '[to] beat', not **bit'* '[to] beat', which is semantically quite plausible).