

Collocations in Translation



Eric Wehrli
LATL - CUI, Université de Genève
Eric.Wehrli@unige.ch



In translation, collocational knowledge is useful

At the parsing level, with respect to disambiguation:

- Categorical ambiguity (*labour costs, budget rules*)
- Attachment ambiguity (*state of emergency, flag of convenience*)

At the transfert level, with respect to

- Word sense disambiguation (cf. examples below)

Word sense ambiguity

standing room *place debout*
significant lead *avance considérable*
loose change *petite monnaie*
close collaboration *collaboration étroite*

The Its-2 translation system

Analysis-Transfer-Generation system based on « deep » linguistics

- source sentence analysis with the Fips parser (Wehrli, 2007, Wehrli & Nerima, 2013), which produces abstract phrase-structure representations, along with argument structures for predicates.
- transfer made on the basis of the phrase-structure representations (recursive traversal, in the order head-left-right), with special transfer rule to handle specific language-pair correspondences (e.g. position of adjectives between English and French, possessive structure, etc.)

black cars → *des voitures noires*

Paul's car → *la voiture de Paul*

- generation involves:
 - transformations of the target structure (tense/mode transfer, movement rules, agreement rules, etc.)
 - morphology (to determine the proper morphological/orthographical form of words)

Collocation translation

- Three steps : identification, transfert, generation
- Collocation identification is achieved by the parser
- Collocation transfer is realized when the lexical transfer of the syntactic head of the collocation is performed.

example: *she put a spell on him* → *elle lui a jeté un sort*

Syntactic head : *put* → *jeter* (vs. *mettre*).

- Target term for the collocate (*spell* → *sort*) is associated with source collocate, to be used upon its lexical transfer.

Collocation translation: several cases

- Source collocation corresponds to a simple target lexeme

avoir envie → *to want*
blind alley → *impasse*
avoir besoin → *to need*
prendre garde → *to watch out*

Transfer head of collocation as target lexeme. Mark collocate as not transferable

- Source lexeme corresponds to target collocation

perjury → *faux témoignage*
to brief → *donner des instructions*
frühstücken → *to have breakfast*

Transfer lexeme as complete collocation constituent

- Source collocation corresponds to target collocation

La nouvelle tombe → *the news break*
To put on a show → *donner un spectacle*
The tax base → *l'assiette fiscale*
To take a look → *jeter un coup d'œil*

Transfer head of collocation as head of target collocation

if local type (noun-noun, adj-noun, etc),

generate complete collocation constituent and mark source collocate as not transferable

else (non-local type such as verb-object)

transfer head of collocation as head of target collocation & link target collocate to source collocate, this information is used when transferring source collocate.

Evaluation

Corpus : Europarl, 3000 sentences

Language pairs : English-French & French-English

Method : BLEU score of corpus is translated **with** and **without** the collocation knowledge

Language pair	With collocations	Without collocations	# sentences affected
English-French	20.19	19.6	506 (17%)
French-English	17.74	17.24	

BLEU scores with and without collocation knowledge

Collocation database

Language	Noun-noun	Adjective noun	Verb object	total
English	5,101	2,627	615	9,490
French	416	5,098	1,280	16,448
German	2,258	458	179	3,237
Italian	127	1,301	257	3,102



UNIVERSITÉ
DE GENÈVE

CENTRE UNIVERSITAIRE
D'INFORMATIQUE