Parsing and Multiword Expressions

THE PARSING CREW

P A R S ≡ M E

July 1, 2016

Matthieu Constant, Miryam de Lhoneux, Verginica Mititelu,
Violeta Seretan, Jakub Waszczuk

Introduction

Motivation: MWEs identification and parsing are tasks that inform each other
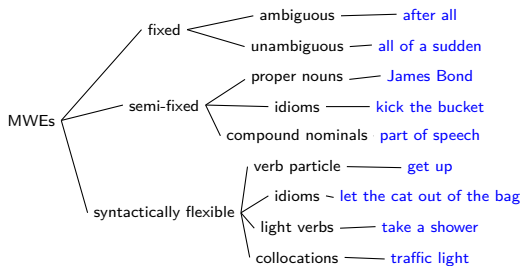
Methods

- Preprocessing: words with spaces
- Word-lattice approach
- Joint/integrated model
- Post-processing

Problem

Problem: no single approach works for all types of MWEs
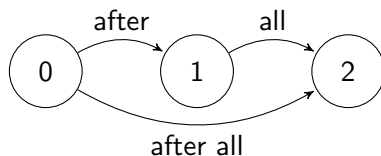Solution: deal with MWE types with approaches adapted to them

MWEs
- fixed
  - ambiguous — after all
  - unambiguous — all of a sudden
- semi-fixed
  - proper nouns — James Bond
  - idioms — kick the bucket
  - compound nominals — part of speech
- syntactically flexible
  - verb particle — get up
  - idioms ~ let the cat out of the bag
  - light verbs — take a shower
  - collocations — traffic light

Methodologies

Pre-processing

Occurrences of MWEs are identified *before* syntactic parsing

Word lattice

- Occurrences of MWE identified *before* syntactic parsing
- Potential ambiguity left for the syntactic parser to resolve

Methodologies

MWE identification integrated into parsing

Occurrences of MWEs identified *during* syntactic parsing

- Symbolic setting: preferring MWE-based analyzes (e.g., when dealing with attachment ambiguities)
- Probabilistic setting: joint-model (syntactic structure and MWEs handled together by a single model)

Post-processing

MWE occurrences identified *after* syntactic parsing (when syntactic structure can help to identify a MWE but not the other way around)

Fixed MWEs

Unambiguous

- Examples: *all of a sudden, ad hoc*
- Solution: word with spaces approach should be optimal

Ambiguous

- Examples: *after all*
  RO: *după aceea – after that.f – 'after that'*
- Solution: lattice / joint model

Semi-fixed MWEs

Compound nouns

- Examples: *credit card*
  RO: *carte de credit – book of credit – 'credit card'*
- Solution: identification in pre-processing => word lattice
- But: ellipsis, coordination (e.g., *car and lorry park*)

Proper nouns

- Examples: *James Bond*, *Grace de Monaco*
- Solution: NER pre-processing => word lattice
- But: knowing syntactic structure can improve NER

Semi-fixed MWEs

Decomposable idioms

- Examples:
  RO: *a-și lua picioarele la spinare*
  to-CL.refl.3.sg.dat. take legs-the at back
  'to run away'

Non-decomposable idioms

- Examples:
  EN: *kick the bucket*
  RO: *a spăla putina*
  to wash barrel-the
  'to run away'

Solution: joint/integrated modeling, post-processing

Syntactically flexible MWEs
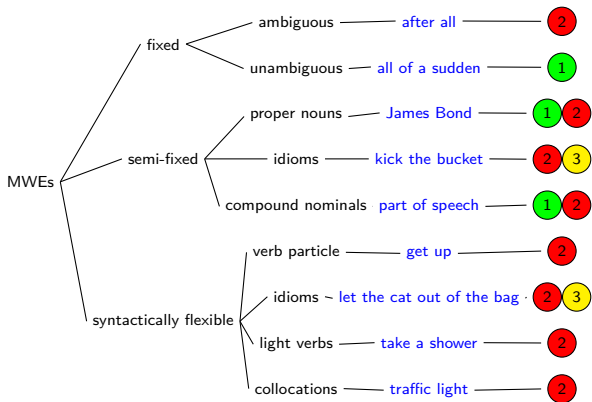
Decomposable idioms

- Examples: *spill the beans*
  RO: *a arunca o privire*
  to throw a.fem look
  'to have a look'

- Solution: joint/integrated modeling, post-processing

Light verb constructions

- Examples: *take a shower*
  RO: *a face față*
  to make face
  'to cope'

- Solution: joint/integrated modeling

Syntactically flexible MWEs

Collocations

- Examples: *traffic light*
- Solution: joint/integrated modeling
- Motivation: can help to deal with attachment ambiguities

MWEs

- fixed
  - ambiguous — after all — **2**
  - unambiguous — all of a sudden — **1**
- semi-fixed
  - proper nouns — James Bond — **1** **2**
  - idioms — kick the bucket — **2** **3**
  - compound nominals — part of speech — **1** **2**
- syntactically flexible
  - verb particle — get up — **2**
  - idioms — let the cat out of the bag — **2** **3**
  - light verbs — take a shower — **2**
  - collocations — traffic light — **2**

1	words-with-spaces or NER preprocessing
2	joint/integrated/lattice
3	post-processing