Dependency Representations for Lexical Segmentation

Matthieu Constant [1,3] Joseph Le Roux [2] Nadi Tomeh [2]

[1]Université Paris-Est, LIGM, CNRS, France

[2]Université Paris 13, LIPN, CNRS, France

[3]Alpage team, INRIA, France

October 20, 2015

This talk

Multiword Expression Identification

- Component of **semantic segmentation** (cf. SemEval 2016 Shared task)
- Processing **running text** (i.e. is not MWE discovery)
- **Framework :** supervised statistical lexicon-based dependency parsing
- **Approaches :** (a) MWE recognizer without syntactic context ; (b) combined with syntactic parser

Contributions

- Exploring use of **dependency representations** for lexical segmentation
- Comparing with use of **sequential representations**
- **Ongoing research :** deep segmentation ; 2-dimensional MWE-aware parser

Lexical segmentation

Definition

- Process that maps a token sequence into a sequence of lexical units
- Lexical units : simple words, multiword expressions (MWE), subpart of tokens (French : *du* $\rightarrow$ *de le*)

Example

- Input : *John made a big deal out of it*
- Output : *John made_a_big_deal out_of it*

Lexical segmentation

Definition

- Process that maps a token sequence into a sequence of lexical units
- Lexical units : simple words, multiword expressions (MWE), subpart of tokens (French : *du* $\rightarrow$ *de le*)

Example

- Input : *John made a big deal out of it*
- Output : *John made_a_big_deal out_of it*

Motivations for Supervised statistical approach

- Learning of **discriminative models** from annotated data
- Can be combined with **external lexicons** to improve MWE coverage
- Help **resolve grammatical ambiguity** given a local context
- Allow to infer some **generalizations** : ex. in French, *coup de N* (strike of N)

MWE Sequential labelling

(Vincze et al. 2011, Constant et al. 2012, Schneider et al. 2014)

Annotating with an IOB-like tagset

I	have	a	bit	of	experience	watching	the	usual	assembly	line
O	B	b	i	o	I	O	O	O	B	I

(example taken from Schneider et al. 2014)

Strength and weakness

- Very accurate and efficient hack in practice
- Theoretically unsatisfactory : bounded embeddings, no interleaved MWEs, no hierachical annotation

Dependency representations

Our idea

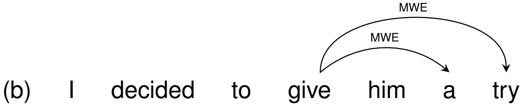
- Representing lexical segmentation with a dependency tree
- Each lexical unit is represented by a subtree, which root is the leftmost token of the lexical unit (not new)
- Similar approaches for word segmentation in Chinese (Zhao et al. 2009 ; Zhang et al. 2014)

Two types of dependencies

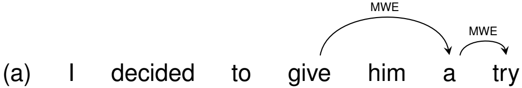
- internal dependencies corresponding to lexical unit subtrees
- external dependencies linking lexical units (subtrees) together

Representation of lexical units

Non-chained representation (Seddah et al 2013)

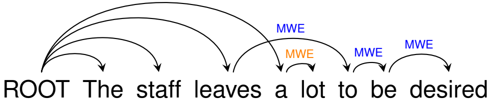


Chained representation (Nivre and Nilsson 2004)

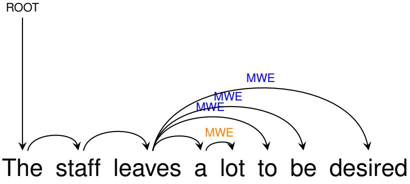


External dependencies

Non-chained representation



Chained representation



Experiment settings

datasets

Language	English		French	Hungarian
Corpus	CMWE (Schneider 14)	Wiki50 (Vincze 11)	FTB (Seddah 13)	Szeged (Vincze 10)
# tokens	55,577	114,335	564,798	1,318,501
# MWEs	3,403	7,490	29,827	3,342
ratio	0.06	0.06	0.05	0.003

Other resources

- parser : TurboParser (Martins el al. 2013) .... but also Mate (Bohnet 10), MaltOptimizer (Ballesteros and Nivre 2013)
- predicted POS tags
- MWE lexicons (morphological features in conll format)

Some preliminary results

CMWE : an "almost" representative case

Annotation	Chained external	Chained internal	Rec.	Prec.	F-score
dependency	-	-	44.9	**65.4**	**53.3**
	-	+	45.1	64.4	53.1
	+	-	43.9	60.1	50.7
	+	+	**45.4**	56.9	50.5
sequential (Schneider et al. 2014)	N/A		**48.3**	61.0	**53.9**

Some preliminary results

CMWE : an "almost" representative case

Annotation	Chained external	Chained internal	Rec.	Prec.	F-score
dependency	-	-	44.9	**65.4**	**53.3**
	-	+	45.1	64.4	53.1
	+	-	43.9	60.1	50.7
	+	+	**45.4**	56.9	50.5
sequential (Schneider et al. 2014)	N/A		**48.3**	61.0	**53.9**

=> Best dep representation : non-chained representation for external dependencies (grand parent feats)

Other datasets

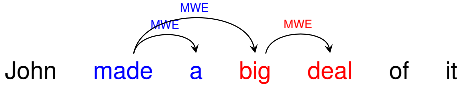
- Wiki50 and Hungarian : dependency > sequential
- FTB : sequential > dependency

Toward a deeper segmentation

Phenomena

- **Embedding :** *John (made a (big deal)) of it*
- **Crossing :** *Luc prend un cachet et demi* (Luc takes one and a half pill)
- **Partial overlapping :** *pay close attention* (Laurent's example)
- **Factorization :** *John and Mary Smith* (general phenomena)

Dependency representations

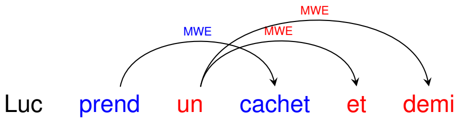


Toward a deeper segmentation

Phenomena

- **Embedding :** *John (made a (big deal)) of it*
- **Crossing :** *Luc prend un cachet et demi* (Luc takes one and a half pill)
- **Partial overlapping :** *pay close attention* (Laurent's example)
- **Factorization :** *John and Mary Smith* (general phenomena)

Dependency representations

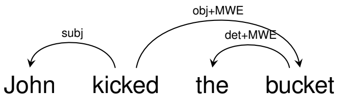


Toward a two-dimensional MWE-parser [under submission]

Related work

- MWE identification combined with syntactic dependency parsing
- **Why ?** both tasks can help each other
- **Joint approach :** a unique parsing model is learned on syntactic treebanks where MWE are integrated as subtrees (Nivre et Nilson 2004, Erygit 2011, Candito et Constant 2014, Nasr et al. 15)
- MWE subtrees can be flat (Seddah et al 2013) or deeper (Vincze et al. 2013)

Starting point : joint MWE and syntactic representation



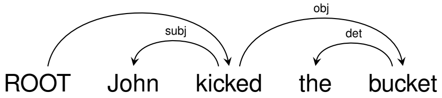
Example inspired by representations in (Vincze et al. 2013, Candito and Constant 2014)

Toward a two-dimensional MWE-aware parser

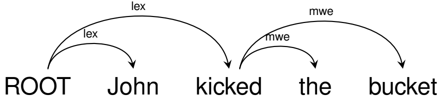
Principle

- **Joint system :** dependency labels concatenate syntactic function and MWE marker
- Why predicting both information exactly at the same time as it increases complexity ?
- **Idea :** duplicate concatenated labels $\rightarrow$ two dimensions (1) syntax, (2) semantic segmentation (e.g. MWE)
- Complex phenomena (embeddings, crossings) can hardly be represented
- **Proposal :** one dimension = one tree ; simultaneously predict both trees

Two-dimensional dependency representation



(a) Syntactic dimension



(b) Semantic segmentation

Links between the two dimensions

- Shared leaves (i.e. words)
- MWE subtree root = MWE syntactic head
- Shared annotation of MWE with irregular syntactic structure
- Extraction of bidimensional features

First results

Experiments

- Implementation in the Easyfirst paradigm (Goldberg and Elhadad 2010)
- Data with shallow MWE annotation : English (Web Treebank), French (FTB)
- Currently, small gains with respect to standard parser... but we're working on it !

Discussions

- **Advantage :** possibility to use a deeper semantic dimension

$\rightarrow$ first tests on Sequoia treebank, deeply reannotated in MWE (very small data)

- **Cons :** Non-factorized representation ; tree is not sufficient (ex. partial overlapping)

Conclusions

- Dependency representation for shallow and deep MWE annotation
- Ongoing integration in a two-dimensional dependency parser