

COST Action IC1207

PARSING and Multi-word Expressions

Towards linguistic precision and computational efficiency
in natural language processing

Working Group 2: PARSING TECHNIQUES FOR MWEs

MWEs in *a* French TAG

PARSEME General Meeting
Athens, 11 March 2014

Introduction

- About *lexicalized* Tree-Adjoining Grammar (TAG):
 - ▶ **Lexicalized** each grammar rule is associated with at least one lexical item (word)
 - ▶ **Tree** a grammar rule corresponds to a (partial) syntactic tree reflecting the syntactical behavior of its lexical item
 - ▶ **Adjoining** a tree can be inserted (adjoined) into another tree
- About the French TAG:
 - ▶ +6,000 syntactic descriptions (built using XMG) [Crabbé, 2005]
 - ▶ Syntax/semantic interface à la Montague [Gardent, 2008]
 - ▶ (syntactic part) freely available on-line (XMG format)
<https://sourcesup.renater.fr/scm/viewvc.php/trunk/METAGRAMMARS/FrenchTAG/?root=xmg>

Introduction

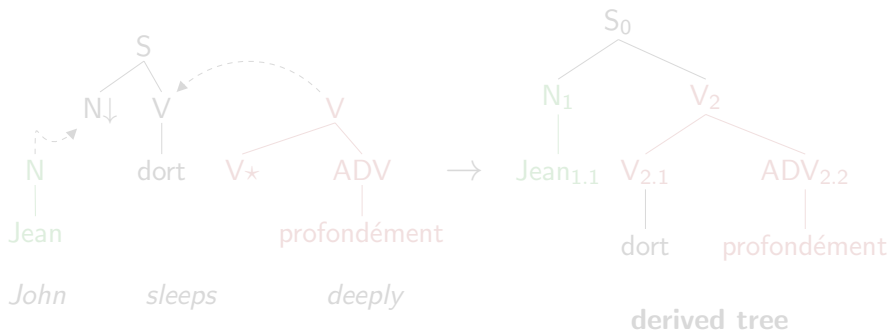
- About *lexicalized* Tree-Adjoining Grammar (TAG):
 - ▶ **Lexicalized** each grammar rule is associated with at least one lexical item (word)
 - ▶ **Tree** a grammar rule corresponds to a (partial) syntactic tree reflecting the syntactical behavior of its lexical item
 - ▶ **Adjoining** a tree can be inserted (adjoined) into another tree
- About the French TAG:
 - ▶ +6,000 syntactic descriptions (built using XMG) [Crabbé, 2005]
 - ▶ Syntax/semantic interface à la Montague [Gardent, 2008]
 - ▶ (syntactic part) freely available on-line (XMG format)
<https://sourcesup.renater.fr/scm/viewvc.php/trunk/METAGRAMMARS/FrenchTAG/?root=xmg>

Plan

- 1 Tree-Adjoining Grammar
- 2 Representing MWEs in TAG
- 3 Conclusion

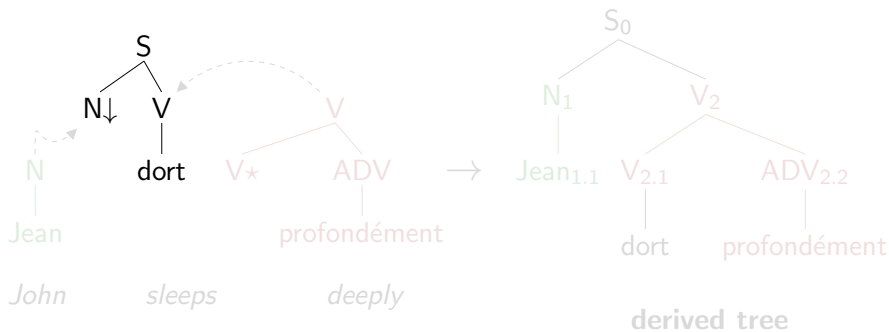
Tree-Adjoining Grammar (1 / 2)

- Tree-rewriting system [Joshi et al., 1975, Joshi and Schabes, 1997]
- Elementary trees built on linguistic well-formedness constraints (lexicalization, predicate/arguments cooccurrence, semantic minimality) [Abeillé, 1993]
- Tree-rewriting operations: **Substitution** / **Adjunction**



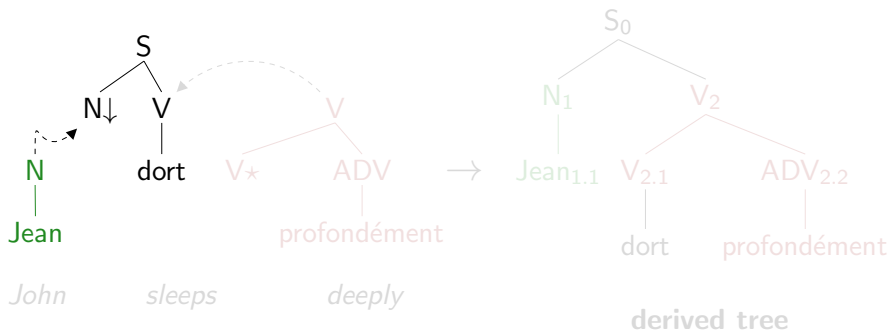
Tree-Adjoining Grammar (1 / 2)

- Tree-rewriting system [Joshi et al., 1975, Joshi and Schabes, 1997]
- Elementary trees built on linguistic well-formedness constraints (lexicalization, predicate/arguments cooccurrence, semantic minimality) [Abeillé, 1993]
- Tree-rewriting operations: **Substitution** / **Adjunction**



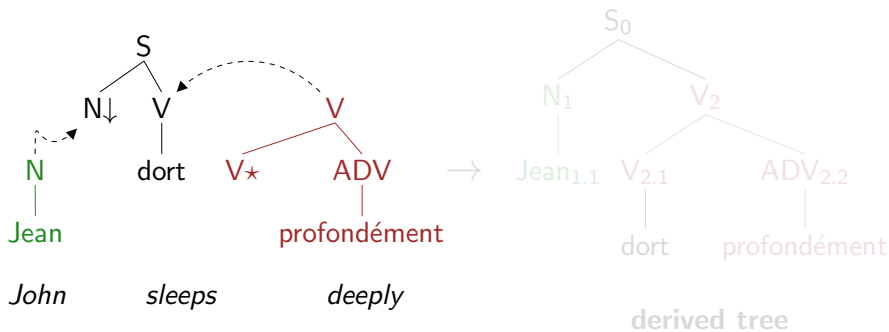
Tree-Adjoining Grammar (1 / 2)

- Tree-rewriting system [Joshi et al., 1975, Joshi and Schabes, 1997]
- Elementary trees built on linguistic well-formedness constraints (lexicalization, predicate/arguments cooccurrence, semantic minimality) [Abeillé, 1993]
- Tree-rewriting operations: **Substitution** / **Adjunction**



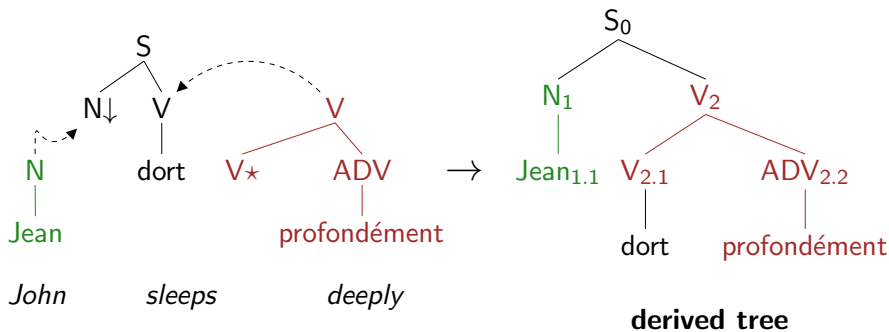
Tree-Adjoining Grammar (1 / 2)

- Tree-rewriting system [Joshi et al., 1975, Joshi and Schabes, 1997]
- Elementary trees built on linguistic well-formedness constraints (lexicalization, predicate/arguments cooccurrence, semantic minimality) [Abeillé, 1993]
- Tree-rewriting operations: **Substitution** / **Adjunction**



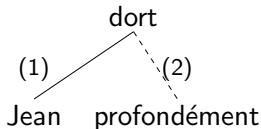
Tree-Adjoining Grammar (1 / 2)

- Tree-rewriting system [Joshi et al., 1975, Joshi and Schabes, 1997]
- Elementary trees built on linguistic well-formedness constraints (lexicalization, predicate/arguments cooccurrence, semantic minimality) [Abeillé, 1993]
- Tree-rewriting operations: **Substitution** / **Adjunction**



Tree-Adjoining Grammar (2 / 2)

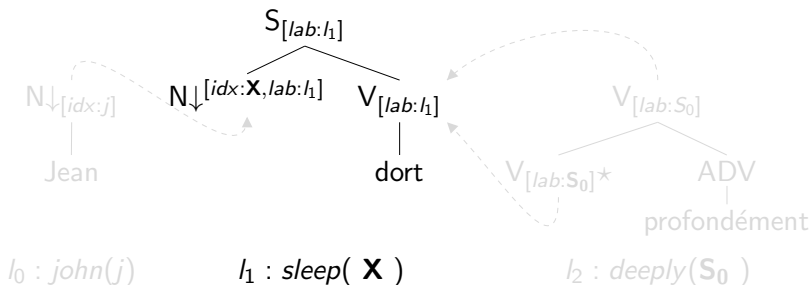
- A TAG derivation produces two structures:
 - ▶ a derived syntactic tree
 - ▶ a derivation tree (log of rewriting operations)
- Example : *Jean dort profondément*



- NB: electronic TAGs are made of 3 components (lexica): tree templates, syntactic entries, morphological entries (see infra)

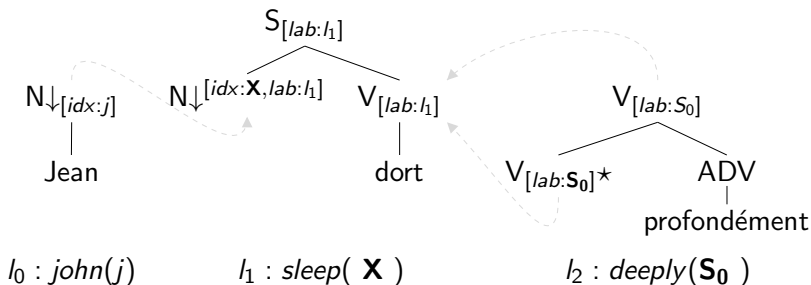
Syntax/semantics interface in FTAG

- Compositional semantics [Gardent and Kallmeyer, 2003]:
 - each elementary tree is associated with a (predicative) **semantic formula**
 - LTAG's feature structures share **unification variables** with semantic formulas



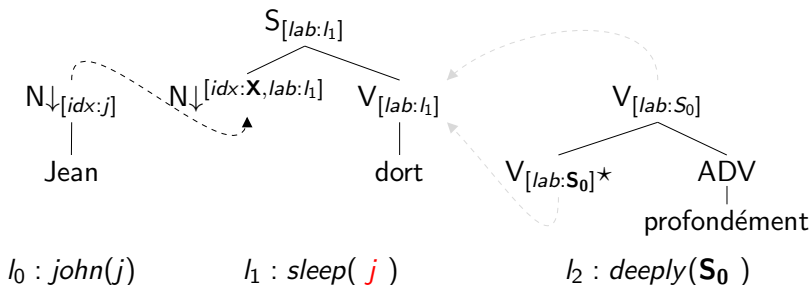
Syntax/semantics interface in FTAG

- Compositional semantics [Gardent and Kallmeyer, 2003]:
 - each elementary tree is associated with a (predicative) **semantic formula**
 - LTAG's feature structures share **unification variables** with semantic formulas



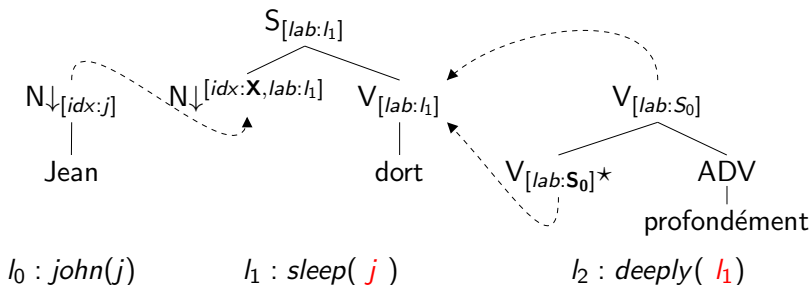
Syntax/semantics interface in FTAG

- Compositional semantics [Gardent and Kallmeyer, 2003]:
 - each elementary tree is associated with a (predicative) **semantic formula**
 - LTAG's feature structures share **unification variables** with semantic formulas



Syntax/semantics interface in FTAG

- Compositional semantics [Gardent and Kallmeyer, 2003]:
 - each elementary tree is associated with a (predicative) **semantic formula**
 - LTAG's feature structures share **unification variables** with semantic formulas



Plan

- 1 Tree-Adjoining Grammar
- 2 Representing MWEs in TAG
- 3 Conclusion

Representing MWEs in TAG

- Very few MWEs in the French TAG, but:
 - ▶ TAG rules provide an *extended domain of locality*
 - ▶ TAG offers a trade-off between expressivity and complexity

- Many attempts for representing MWEs within TAG^(*):

[Abeillé, 1988, Abeillé and Schabes, 1989, Abeillé, 1995, Schuler and Joshi, 2011, Green et al., 2011]

(*) or tree-based grammars

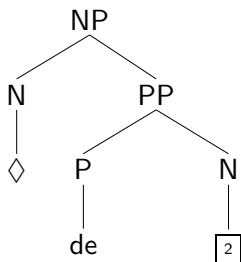
Representing MWEs in TAG

- Very few MWEs in the French TAG, but:
 - ▶ TAG rules provide an *extended domain of locality*
 - ▶ TAG offers a trade-off between expressivity and complexity
 - Many attempts for representing MWEs within TAG^(*):
[Abeillé, 1988, Abeillé and Schabes, 1989, Abeillé, 1995, Schuler and Joshi, 2011, Green et al., 2011]
- (*) or tree-based grammars

Representing MWEs in TAG (continued)

Example 1: *tour de* { *magie* | *passe-passe* | *force* }, *pomme de terre*

(α)



Tree template

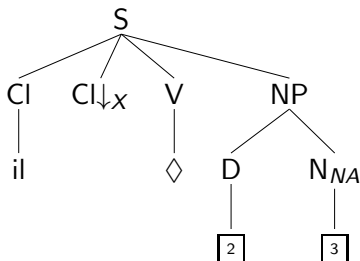
HEAD	tour
CAT	N
MORPH	$\begin{bmatrix} \text{gen} & \text{masc} \\ \text{num} & \text{sg} \end{bmatrix}$
CO-ANCHOR	$\begin{bmatrix} \boxed{2} & \text{magie} \end{bmatrix}$
FAMILY	α
SEM	<i>trick(t)</i>

Lexical entry

Representing MWEs in TAG (continued)

Example 2: *il lui pousse (souvent) des (*grandes) ailes*
 it him grows (often) (*big) wings
 he (often) spreads his (*big) wings

(β)



Tree template

HEAD	pousse				
CAT	V				
MORPH	<table border="1"> <tr> <td>pers</td> <td>3rd</td> </tr> <tr> <td>num</td> <td>sg</td> </tr> </table>	pers	3rd	num	sg
pers	3rd				
num	sg				
CO-ANCHOR	<table border="1"> <tr> <td>2</td> <td>des</td> </tr> <tr> <td>3</td> <td>ailes</td> </tr> </table>	2	des	3	ailes
2	des				
3	ailes				
FAMILY	β				
SEM	<i>enhance_skills(X)</i>				

Lexical entry

Representing MWEs in TAG (continued)

Advantages :

- TAG allows to express long-distance dependencies within single elementary rules
- Null-adjunction constraints can prevent modifications on given syntactic nodes

Drawbacks :

- Various syntactic structures of MWEs need to be enumerated
- High redundancy between structures
- Expression of semantic constraints still unclear

Representing MWEs in TAG (continued)

Advantages :

- TAG allows to express long-distance dependencies within single elementary rules
- Null-adjunction constraints can prevent modifications on given syntactic nodes

Drawbacks :

- Various syntactic structures of MWEs need to be enumerated
- High redundancy between structures
- Expression of semantic constraints still unclear




Conclusion

- lexicalized TAG can encode various MWEs at the price of structural redundancy
 - metagrammar (XMG) [Crabbé et al., 2013]
- price to pay at parsing as well
 - lexical disambiguation (supertagging) [Gardent et al., 2014]
- Future work :
 - define abstractions over MWEs' structures
 - use MWE-recognition to improve supertagging
 - use semantic information to reduce search space at parsing


Conclusion

- lexicalized TAG can encode various MWEs at the price of structural redundancy
 - metagrammar (XMG) [Crabbé et al., 2013]
- price to pay at parsing as well
 - lexical disambiguation (supertagging) [Gardent et al., 2014]
- Future work :
 - define abstractions over MWEs' structures
 - use MWE-recognition to improve supertagging
 - use semantic information to reduce search space at parsing



References I

-  Abeillé, A. (1988).
Parsing French with Tree Adjoining Grammar: some linguistic accounts.
In COLING-88: Proceedings of the 12th International Conference on Computational Linguistics, volume 1, pages 7–12, Budapest, Hungary.
-  Abeillé, A. (1993).
Les Nouvelles Syntaxes.
A. Colin – Paris.
-  Abeillé, A. (1995).
The Flexibility of French Idioms: A Representation with Lexicalised Tree Adjoining Grammar.
In Everaert, M., van der Linden, E.-J., Schenk, A., and Schreuder, R., editors, *Idioms: Structural and Psychological Perspectives*, chapter 1. Lawrence Erlbaum Associates.




References II

-  Abeillé, A. and Schabes, Y. (1989).
Parsing Idioms in Lexicalized TAGs.
In *Proc. of the 4th EACL*, pages 1–9, Manchester, UK.
-  Crabbé, B. (2005).
Grammatical development with XMG.
In *Proceedings of the 5th International Conference on the Logical Aspects of Computational Linguistics (LACL05)*, Bordeaux, France.
-  Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., and Parmentier, Y. (2013).
XMG : eXtensible MetaGrammar.
Computational Linguistics, 39(3):591–629.



References III

-  Gardent, C. (2008).
Integrating a unification-based semantics in a large scale Lexicalised Tree Adjoining Grammar for French.
In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08)*, pages 249–256, Manchester, UK.
-  Gardent, C. and Kallmeyer, L. (2003).
Semantic Construction in FTAG.
In *EACL 2003, 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 123–130, Budapest, Hungary.

References IV

-  Gardent, C., Parmentier, Y., Perrier, G., and Schmitz, S. (2014).
Lexical Disambiguation in LTAG using Left Context.
In *Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers.* Springer.
-  Green, S., de Marneffe, M.-C., Bauer, J., and Manning, C. D. (2011).
Multiword Expression Identification with Tree Substitution Grammars: A Parsing *tour de force* with French.
In *EMNLP*, pages 725–735.
-  Joshi, A. K., Levy, L. S., and Takahashi, M. (1975).
Tree adjunct grammars.
Journal of the Computer and System Sciences, 10:136–163.

References V

-  Joshi, A. K. and Schabes, Y. (1997).
Tree adjoining grammars.
In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*. Springer Verlag, Berlin.
-  Schuler, W. and Joshi, A. K. (2011).
Tree-rewriting Models of Multi-word Expressions.
In *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World, MWE '11*, pages 25–30, Stroudsburg, PA, USA. Association for Computational Linguistics.