Enhanced UD dependencies with Neutralized Diathesis Alternations

Marie Candito[1], Bruno Guillaume[2],
Guy Perrier[3] and Djamé Seddah[4]

[1]Univ Paris Diderot, [2]Loria,
[3]Univ de Lorraine, [4]Univ Paris Sorbonne

Introduction

- UD scheme favors dependencies between content words

  - better cross-linguistic generalization

  - more semantic-oriented dependencies

- Yet, UD dependencies remain syntactic trees

  - Pb for well-known syntactic/semantic mismatches

Syntactic/Semantic mismatches

- Argument sharing

  - control verbs, Right-node raising, coordination…

- 1 syntactic argument = no semantic argument

  - e.g. impersonal construction

    FR: il est arrivé 3 personnes
    it is arrived 3 people
    « 3 people arrived »

- 2 syntactic arguments = 1 semantic argument

  - e.g. raising verbs, predicative complements

    FR: Marie a trouvé Anna fatiguée
    Marie has found Anna tired
    « Marie found that Anna was tired »

Beyond dependency trees

- Many proposals towards predicate-argument structures

  - Stanford dependencies (de Marneffe and Manning 08)

  - Graph banks

    - cf. in-depth analysis of 4 English graph-banks by Kuhlman & Oepen (CL, 2016)

    - the Semeval 2014 shared task on « broad coverage semantic dependency parsing » (Oepen et al. 14)

  - « Deep syntax »

    - Spanish: MTT deep trees (Ballesteros et al. 16)

    - French: Deep syntactic graphs (Candito et al. 14)

  - Tectogrammatical structures in Prague Dependency treebank …

More or less semantics

- In these proposals, e.g. labels are more or less semantic-oriented

  - syntactic labels

  - numbered arguments

    - arg0, arg1, arg2 …

    - MTT : deep syntactic arguments I, II, III …

  - semantic roles

    - patient, addressee, beneficiary …

    - as in tectogrammatical structures in Prague DT

Enhanced UD graphs

- « Enhanced dependencies »

  - Enhanced / enhanced**++** for English (Schuster & Manning, 16)

  - proposed as optional in UD v2.0

  - available for a few languages (Russian, Finnish)

Enhanced UD graphs

- 5 enhancements

  - subj. of infinitives in control/raising constructions

    Paul seems to run: **run —nsubj—> Paul**

  - propagation of conjuncts

  - antecedent of relative pronouns

  - markers as suffixes in labels

   **went —obl:into—> house**

  - null nodes for elided predicates

    Mary wants to buy a book and Jenny **N1 N2** *a CD*

This work

- Yet another proposal for enhanced UD: « **Enhanced-diat** »

    - that neutralizes syntactic alternations

- Implemented and evaluated on French

Enhanced-diat

- Enhanced-diat graphs remain mostly syntactic
  - in particular, we keep **UD syntactic labels**
  - as starting point for various kinds of semantic representations

```
Syntactic tree
      |
      v
Deep syntactic graph
   /    |    |    \        \
  v     v    v     v        v
PAS   AMR  MRS    ...
```
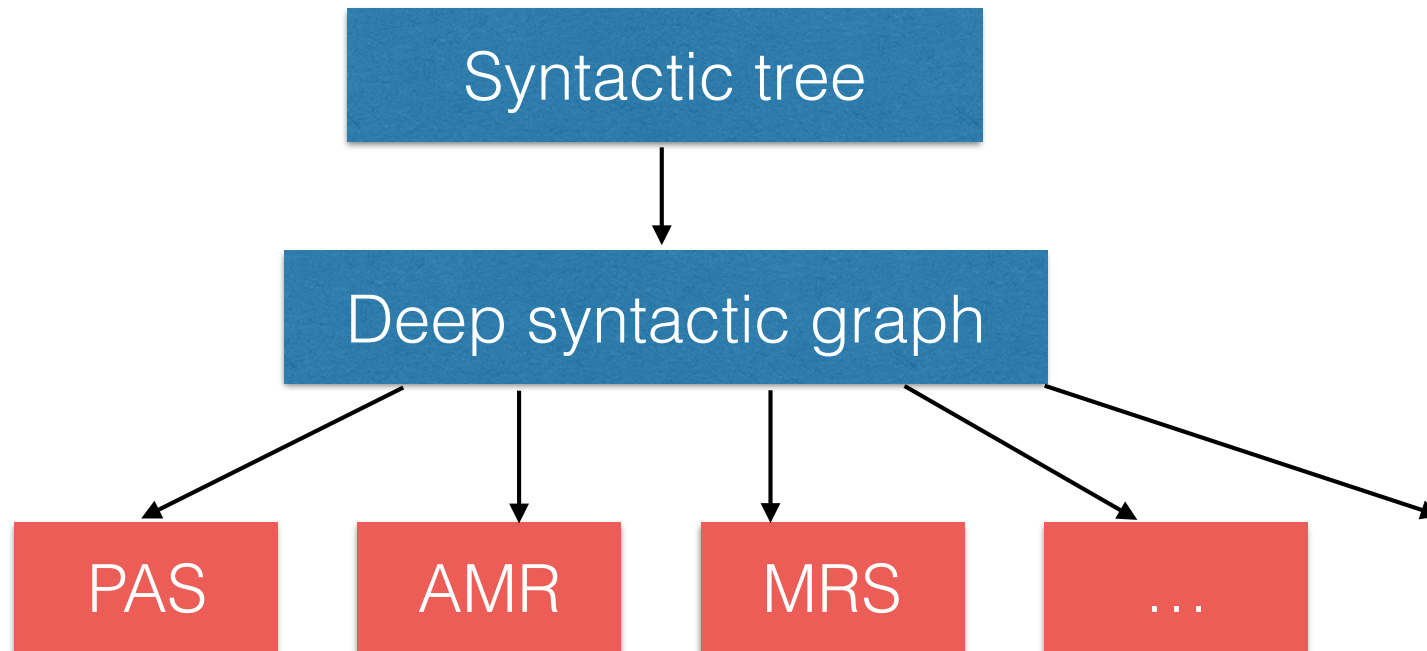
Enhanced-diat

- 2 enhancements over enhanced UD:

  - Add even more argumental edges, either

    - some fully determined by syntax:

      - control nouns, adj, some participles, gerunds

    - other cases not fully determined but most frequent

  - Neutralize syntactic alternations

    - recover canonical subcat frame

More argumental edges: Example: noun-modifying participle

(a)

(b)
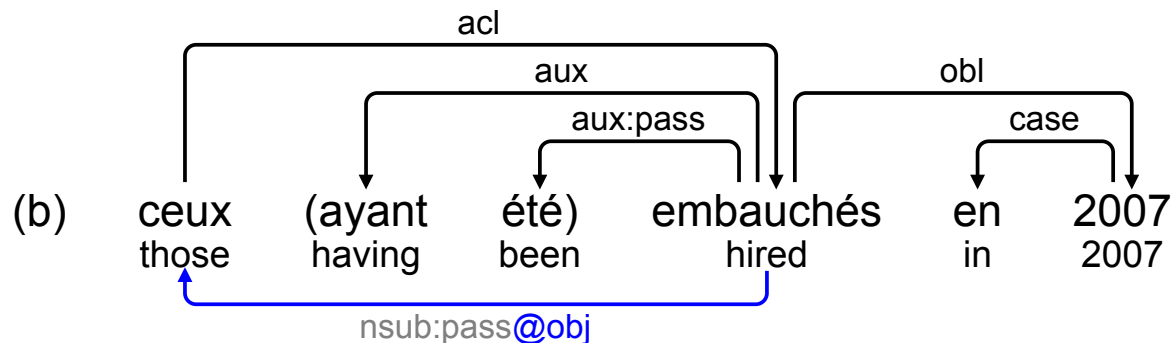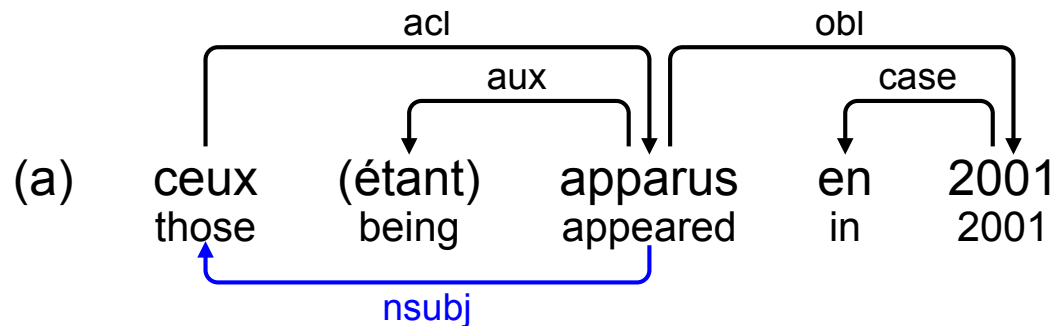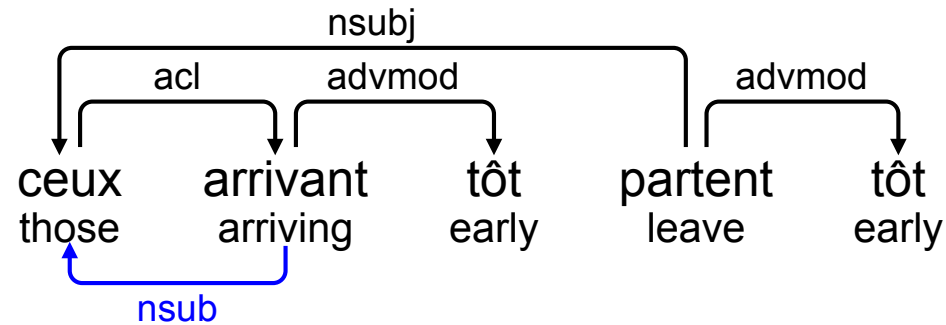
More argumental edges:
Example: infinitive adverbial clauses

- When main verb is active, with non expl subject

- **subject of infinitive = subject of main verb**

- in most cases (83% on Sequoia corpus)

  Il mangera avant de jouer

  He will-eat before to play

  « He will eat before playing »

- counter-example:

  D'autres photos ont subi des retouches pour accentuer le drame

  Other photos have undergone modifications to accentuate the drama

Neutralizing syntactic alternations

* recover « canonical » grammatical functions

  * the function you would get in active personal voice

* cheap way to limit **linking diversity**

  * e.g. proved useful for FrameNet parsing (Michalon et al. 16)

* massive for passive

* other cases (see paper):

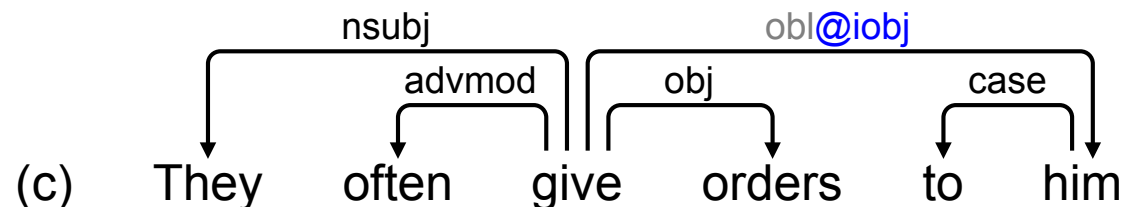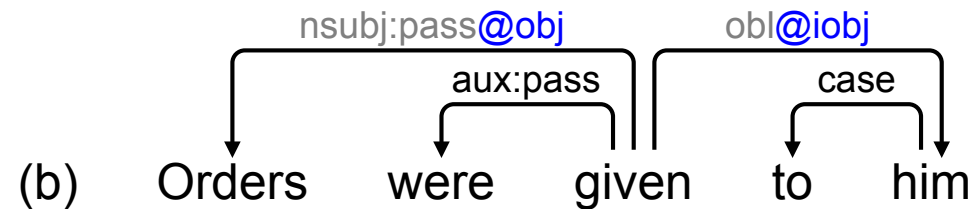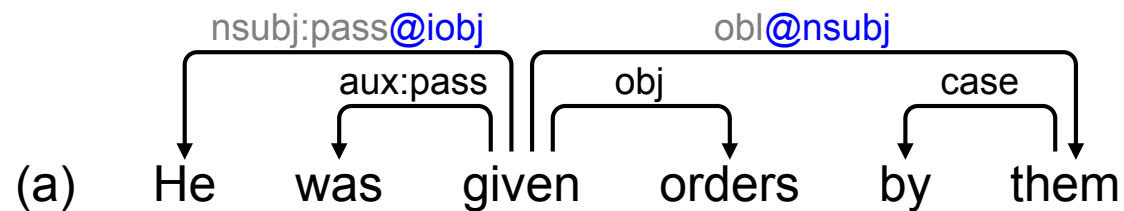 * impersonal, causative, mediopassive

Neutralizing syntactic alternations
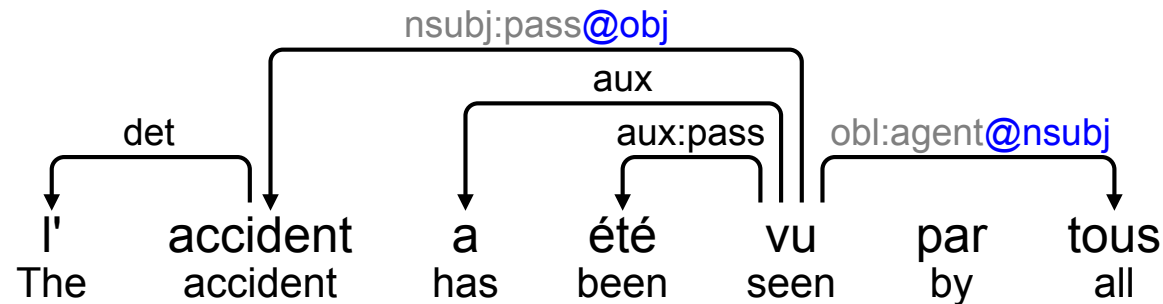
- Note:

  - nsubj:pass / csubj:pass not enough to recover all arguments of passive (obl / obl:agent)

  - UD choice to distinguish functions according to POS of dependent (nsubj/csubj, obj/xcomp…) augments linking diversity

Syntactic alternation normalization for English ditransitives

- Take canonical subcat :

  - They$_{(nsubj)}$ gave him$_{(iobj)}$ orders$_{(obj)}$

(a) nsubj:pass@iobj, aux:pass, obl@nsubj, obj, case: He was given orders by them

(b) nsubj:pass@obj, aux:pass, obl@iobj, case: Orders were given to him

(c) nsubj, advmod, obl@iobj, obj, case: They often give orders to him

Obtaining enhanced-diat graphs for French

- 2 teams, 2 graph-rewriting systems

  - GREW (Guillaume et al. 12) : 157 rules

  - OGRE (Ribeyre et al. 12) : 115 rules

  - building on rules written for producing deep-sequoia (Candito et al. 14; Perrier et al. 14)

- rules written supposing gold surface tree

- mix of

  - purely deterministic cases (e.g. control verbs)

  - cases previously analyzed as « almost deterministic »

    - cf. previous example of infinitive adverbial clauses

Gold corpus for evaluation

- We produced gold graphs for 200 sentences

  - 100 from UD_French

  - 100 from UD_French-Sequoia

  - bias: obtained through adjudication of the 2 rule-based systems outputs

Quantitative assessment of enhancements

- **4804** edges in the 200 sentence gold corpus

- **956** are **argumental dependents of verbs**

  - approximated using core argument labels (nsubj,csubj,obj,iobj,ccomp,xcomp) + obl label

- edges added (set N): **18.9 %**

- edges with neutralized label (set A) : **13,9 %**

- N U A represent **26.7 %** of arguments of verbs

Evaluation in 2 modes

- **PA+** : with manual pre-annotation of certain phenomena

  - expletive « il »

  - reflexive clitic « se » status (for mediopassive)

  - canonical subjects in causative constructions

  - agents of passives (by-phrases : obl:agent)

- **PA-** : no pre-annotation, handling by rules known to be approximative

Evaluation in 2 modes

		PA−		PA+	
		SEQ$_{test}$	UD$_{test}$	SEQ$_{test}$	UD$_{test}$
All	OGRE	98.81	99.17	99.46	99.40
edges	GREW	99.44	99.54	99.69	99.66
$N \cup A$	OGRE	86.20	89.89	92.51	91.71
edges	GREW	93.42	94.31	95.77	95.39

Table 1: Evaluation of rule-based systems producing enhanced graphs: F-measures computed on all edges (top) or only on edges in N or A (bottom);

Conclusion

- Production of high quality enhanced UD graphs proved feasible for French

  - a little better with pre-annotation of a few not-so-deterministic phenomena

- **Quality**: accurate enough to serve as pseudo-gold for data-driven methods

- **Impact**: when considering arguments of verbs:

  - 19% are enhanced edges

  - 14% have a label modified by neutralizing syntactic alternation

Conclusion (cont)

- Other languages ?

  - Romance

  - English:

    - diathesis alternations used for some experiments for the EPE shared task

    - Paris / Stanford system (Schuster et al. 17)

Thank you!
Questions?

data / rules available at
https://github.com/bguil/Depling2017