# Because Syntax Does Matter: Improving Predicate-Argument Structures Parsing with Syntactic Features The Case of French

**Corentin Ribeyre**[⋆] Éric de la Clergerie[°] Djamé Seddah[°◇]
Marie Candito[°,†]

[⋆]Université de Genève [°]Alpage - Inria [◇]Université Paris Sorbonne
[†]Université Denis Diderot

1st Paris NLP Meetup

# Context

## Who am I ?

- Associate Professor (MdC) at Paris Sorbonne (Paris IV)
- Researcher in the Inria's Alpage project team
- Focusing in robust parsing, user generated content, morphologically-rich languages and syntax to semantic interface.

## This work

Based on Corentin Ribeyre's Phd Thesis and side projects from the team (notably Marie Candito, Eric de la Clergerie, ,..)

# ALPAGE ?

## A multidisciplinary Inria-University team

- linguistics researchers
- Computer science researchers
- definitly oriented towards Natural Language Processing

## Long standing tradition of partnerships with the industry

- CIFRE PhD thesis (Viavoo, Vera, AFP,..)
- Collaboration contracts (Kwaga, Proxem, Davi,..)
- ANR Projects or FUI (Mandriva -RIP-, Thales,..)

# Example: Information Extraction with the AFP

# What's Natural Language Processing?

## NLP aims at structuring language productions

- in minimal sense unit : words, morphemes..
- in syntactic unit/relation : subject, verb, object, modifier
- in semantic unit: who did what to whom? who did say what?

## This structuring implies the definition of these units as well as their scopes

- "word" vs token: ***chépa***, ***'la*** *pas* [*cassé sa pipe*] *lui deja, wsh?*

⇒ *Typographic segmentation doesn't hold*

- regular vs non-canonical syntax: *John is tired* vs *dunno dude too tired to think 2day*

⇒ *Who is tired? the speaker or someone else?*

- The context of a production: *I don't feel that brand and stuff.*

⇒ *What brand? what stuff? who is he answering to?*

# NLP: HOW DOES IT WORK?

## Using linguistics knowledge. One principle, two schools:

- **(i)** Building grammars, extraction rules and associated software.

⇒ **Old-school** *approach, coslty. Precise but very application-dependant.*

- **(ii)** Building annotated data and let learning models that will do the same as (1) (but better, certainly faster)

⇒ **Data-driven** *approach, we focus on the model that can generalize the data. Flexible but domain sensitive, (relatively) cheap*

# NLP: HOW DOES IT WORK?

**Using linguistics knowledge. One principle, two schools:**

- **(i)** Building grammars, extraction rules and associated software.

⇒ **Old-school** *approach, coslty. Precise but very application-dependant.*

- **(ii)** Building annotated data and let learning models that will do the same as (1) (but better, certainly faster)

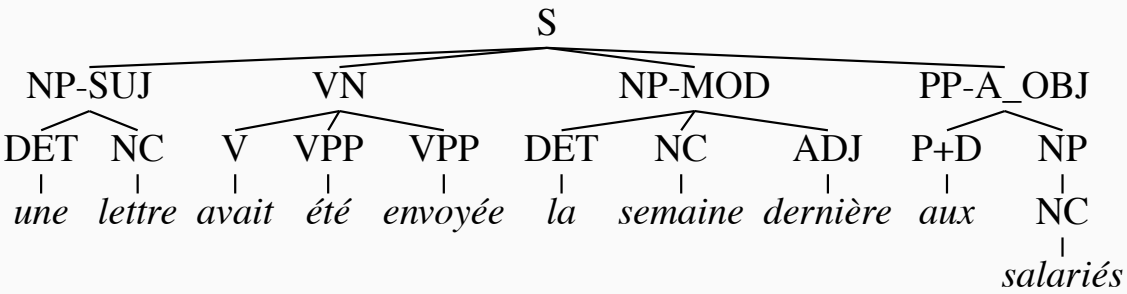
⇒ **Data-driven** *approach, we focus on the model that can generalize the data. Flexible but domain sensitive, (relatively) cheap*

- **(!)** Building nothing and count on massive quantity of data to detect regularities, bring out information

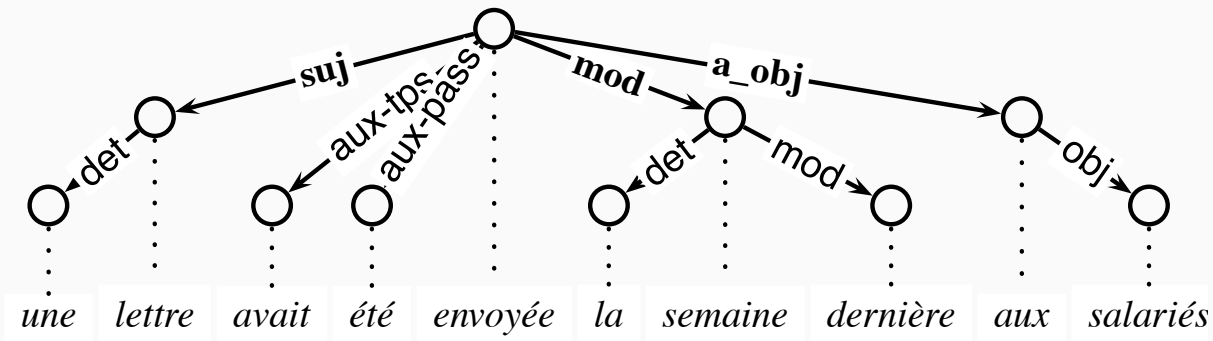
⇒ **non-supervised** *approach (=no prior linguistics knowledge), need much more data (but cheaper). Text-mining ≠ NLP.*

# What do Linguistics data look like?

## A constituent tree (bracketed format: *(SENT (NP-SBJ ..) (VN (V had) ..*




## Dependency tree(tabular format: *csv*)



# And?

## Treebanks, data set and evaluation

- a set of annotated parse trees (dep. or const.)is called a **treebank**.
- Set of linguistics decisions is called an **annotation scheme** (many exists, very hard to design).
- The task of predicting such structures is called **parsing**
- Evaluation is done on **comparing predicted trees vs gold ones**
- **Different metrics** based on the structures itself. (percentage of matching subtrees (F-score), percentage of correct predictions token by token (Accuracy) , etc.)

# Introduction

## The on-going trend that hides the forest

- For years now, data-driven syntactic parsing has reached good performances.
  - Around 92% (LAS) on English
  - Between 85% - 88% (LAS) on morphologically richer languages (French, German, Korean, Arabic, . . . ).
- Trouble is that these parsers only focus on surface syntax with various levels (often limited) of non-projectivity.

# Introduction

## The on-going trend that hides the forest

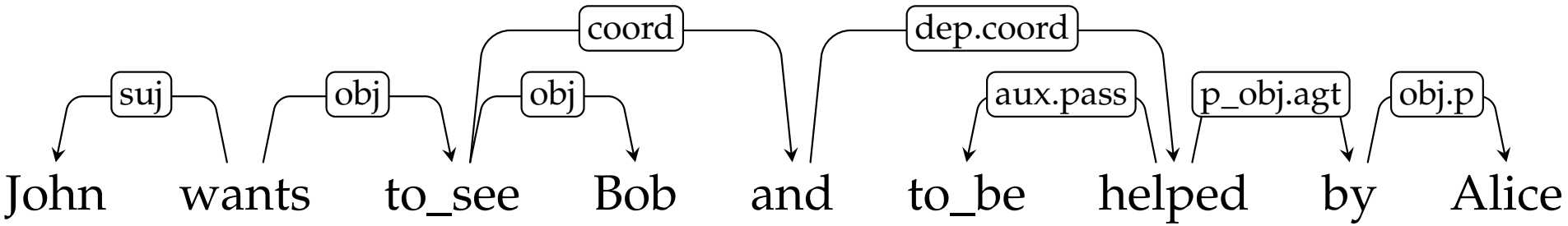
- For years now, data-driven syntactic parsing has reached good performances.
  - Around 92% (LAS) on English
  - Between 85% - 88% (LAS) on morphologically richer languages (French, German, Korean, Arabic, . . . ).
- Trouble is that these parsers only focus on surface syntax with various levels (often limited) of non-projectivity.
- For downstream applications relying on further semantic processing, full argument structures are needed

# Introduction

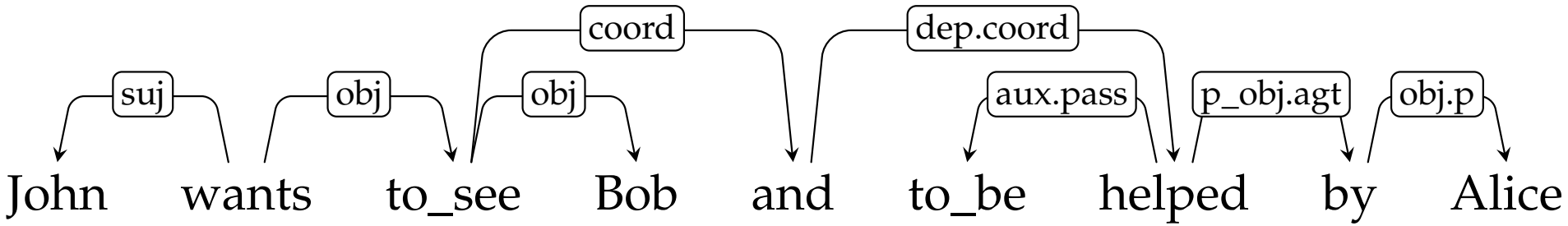
## The on-going trend that hides the forest

- For years now, data-driven syntactic parsing has reached good performances.
  - Around 92% (LAS) on English
  - Between 85% - 88% (LAS) on morphologically richer languages (French, German, Korean, Arabic, . . . ).
- Trouble is that these parsers only focus on surface syntax with various levels (often limited) of non-projectivity.
- For downstream applications relying on further semantic processing, full argument structures are needed
- In other words “what is the subject of that causative?”

# Beyond Surface Syntax



# Beyond Surface Syntax

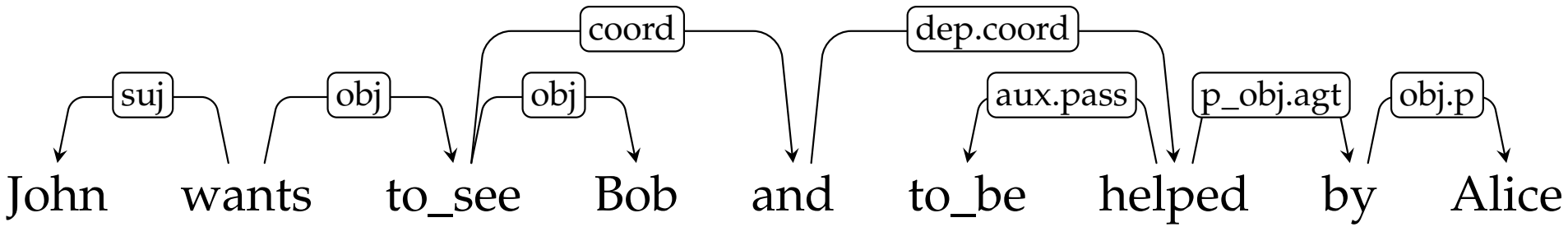



- Some informations are not expressed at this level but they are needed for semantic applications.

⇒ So, such a tree is called a **surface syntactic tree**.

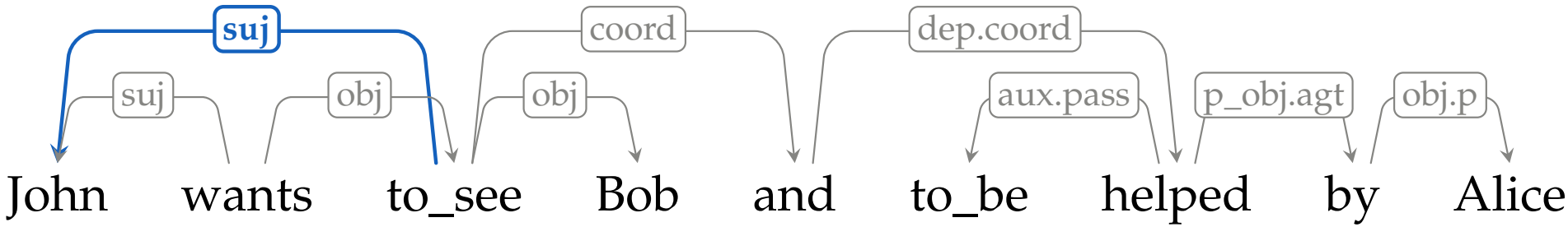
# Beyond Surface Syntax

**Toward a deeper structure**



# Beyond Surface Syntax

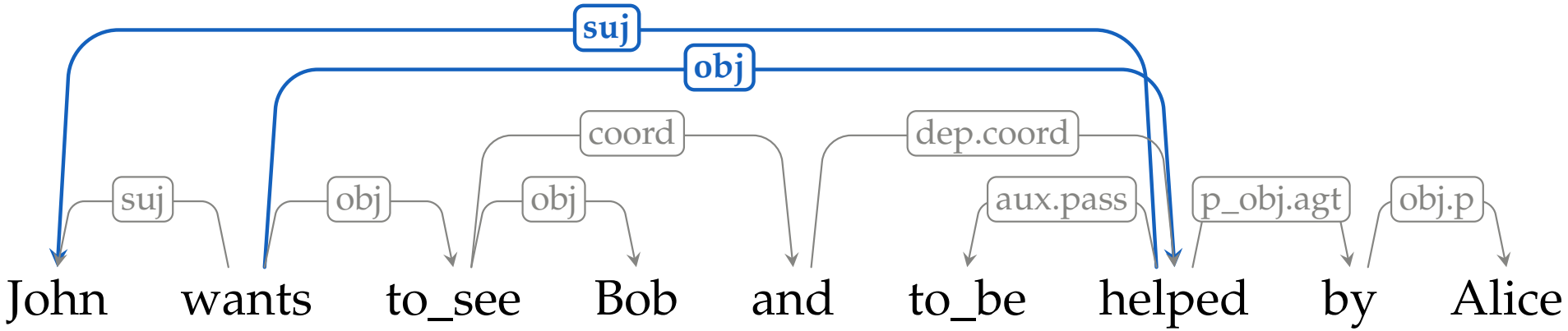
**Toward a deeper structure**




1. Across many phenomena, arguments are not expressed
   - Infinitives without *realized* subjects (controlled subjects, causative).

# Beyond Surface Syntax

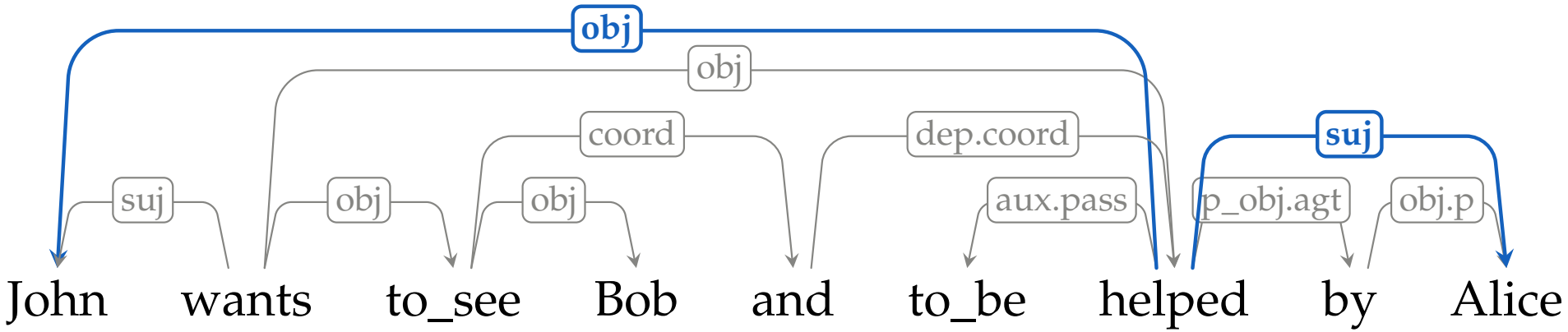
## Toward a deeper structure




1. Across many phenomena, arguments are not expressed
   - Infinitives without *realized* subjects (controlled subjects, causative).
   - Subject ellipsis in coordinations.

# Beyond Surface Syntax

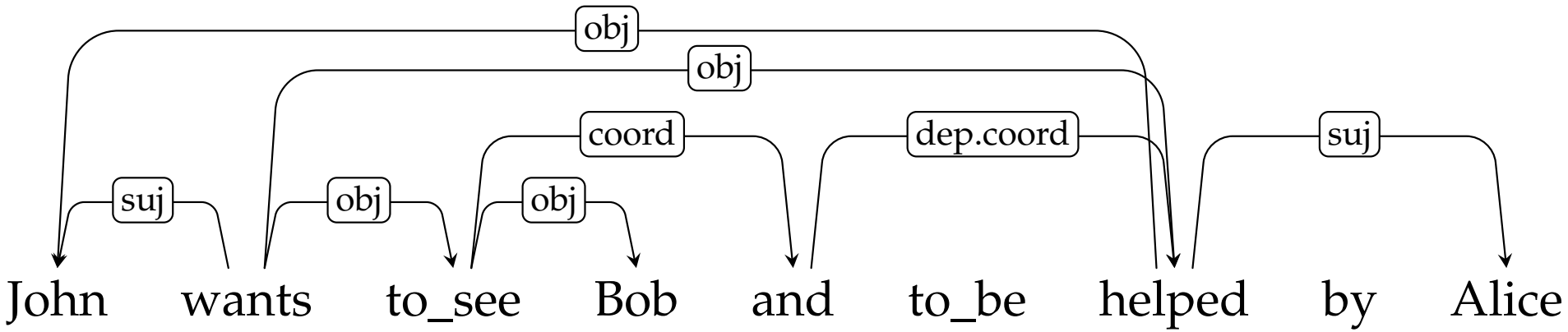
**Toward a deeper structure**




1. Across many phenomena, arguments are not expressed
2. Arguments need to be stable
   - Regardless of diathesis change (syntactic alternations)

# Beyond Surface Syntax

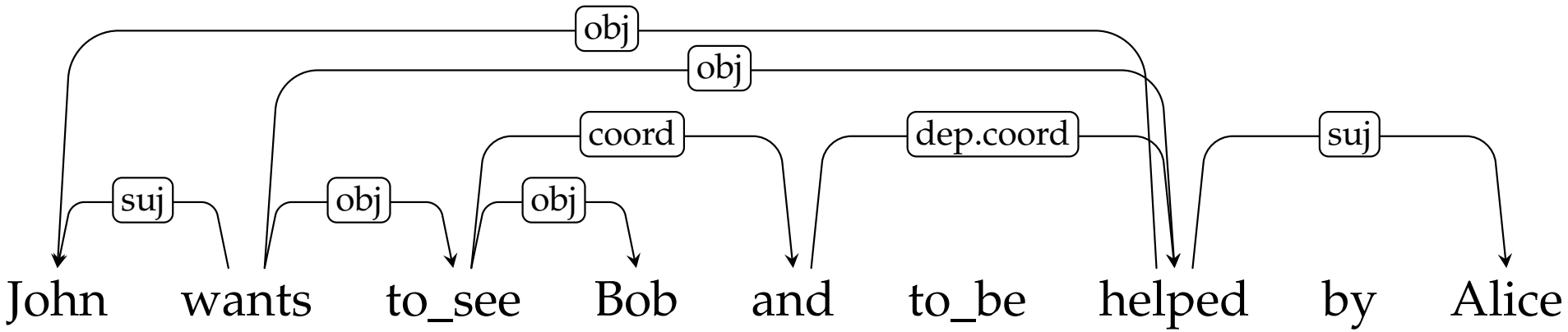
**Toward a deeper structure**




1. Across many phenomena, arguments are not expressed
2. Arguments need to be stable
3. Discarding semantically empty words.

# Beyond Surface Syntax

**Toward a deeper structure**




1. Across many phenomena, arguments are not expressed
2. Arguments need to be stable
3. Discarding semantically empty words.

⇒ This representation is a **deep syntactic graph**

# NOT SO NEW..

## 10-15 years ago

- Rise of treebank-based wide coverage deep syntax parsers

⇒ *LFG (Cahill et al, 2004), HPSG (Miyao and Tsuji, 2005), CCG (Hockenmeyer and Steedman, 2002)*

- Based on costly efforts to rewrite treebank wrt a given theory
- in most cases, the parser was tied to its training data

## 2006-2007: The Dependency Revolution

- the ConLL shared tasks
- multilingual data by essence, surface syntax tree by nature
- pure data driven method: everyone could play!
- dominant framework of choice (fast and easy to process)

# NOT SO NEW..

## 10-15 years ago

- Rise of treebank-based wide coverage deep syntax parsers

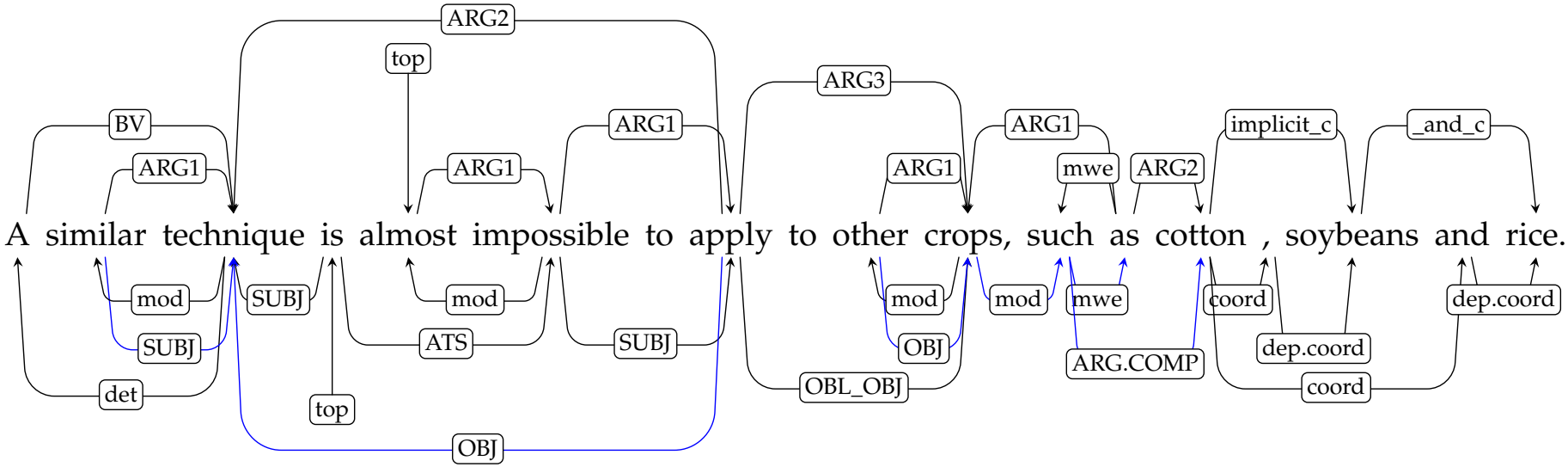
⇒ *LFG (Cahill et al, 2004), HPSG (Miyao and Tsuji, 2005), CCG (Hockenmeyer and Steedman, 2002)*

- Based on costly efforts to rewrite treebank wrt a given theory
- in most cases, the parser was tied to its training data

## 2014-2015: The Semantic Graph-Parsing Evolution

- Semeval's Broad coverage semantic parsing shared tasks (Oepen et al, 2014,2015)
- bring to light new data sets, graph-based, deep syntax/semantic analysis
- monolingual at first (eg. DM, PAS, ..), plus Chinese and Czech in 2015.
- Spanish in 2014 (Ballesteros et al), French in 2014 (Ribeyre et al)

# What do these Data Set Look like? : DM vs DeepFTB




- top: DM treebank derived from the DeepBank (Flickinger et al, 2012)
- bottom: Deep-Ftb (FTB + semi-auto rule-based conversion to deep-syntax scheme (Candito et al, 2014))

# Treebank Properties

| | DM | | Deep-ftb | |
|---|---|---|---|---|
| | Train | Dev | Train | Dev |
| # Sentences | 32k | 1.6k | 14.7k | 1.2k |
| # Tokens | 742k | 36k | 457k | 40k |
| % Void tokens | **21.63** | **21.58** | 11.97 | 12.19 |
| % Void tokens (no punct.) | NA | NA | 35.34 | 35.57 |
| # Edges | 559k | 27k | 424k | 37k |
| % Crossing edges | 4.24 | 4.05 | 3.70 | 3.87 |
| Edges/sentences | 17.29 | 17,21 | 29.14 | 30.05 |
| Label set | 52 | 36 | 27 | 24 |

- Both corpora are comparable in term of semantically empty tokens
- The Deep-FTB has more edges per sentences –> syntactically denser. Twice as much labels for DM (cf. coordination)

# This Work

## Deep syntax parsing is being addressed

- English: DM parsing performance crosses 89.5-7 LF (Du et al, 2014, Ribeyre et al, 2015, Almeida and Martins, 2016),
- For Czech and Chinese, results are lower (especially Czech) see (Oepen et al, 2015)

## How far we can go in Deep-syntax parsing of French?

1. Is this new data set self-sufficient? (is it parsable?)
2. Are these new annotated phenomena (eg. subject ellipsis, LDDs) that hard to parse?
3. Does adding more syntactic context help?
4. Can we start working on the semantic side?

# Experiment Protocol

## High order parsing models

Extended version of the TurboParser for parsing general graphs (Martins and Almeida, 2014).

- Dual decomposition arc-factored model.
- One of the top performers of the SemEval 2014 shared task.

→ We extended the feature capabilities of the parser which were heavily restricted (Ribeyre et al, 2015).

## Realistic Scenario

- Predicted POS and morphological features (including mwe predictions for French (SPMRL Shared task 2014 FTB)

# Experiment Protocol

## High order parsing models

Extended version of the TurboParser for parsing general graphs (Martins and Almeida, 2014).

- Dual decomposition arc-factored model.
- One of the top performers of the SemEval 2014 shared task.

→ We extended the feature capabilities of the parser which were heavily restricted (Ribeyre et al, 2015).

## Baseline system

- rule based conversion (same as used for creating the DeepFTB)
- TurboParser (surface dependency, (Martins et al., 2010))
- FTB perf: dev: 80.86 LF, test: 80.45

# Baseline Results

| Dev Set | DM | DeepFTB |
|---|---|---|
| Baseline TSParser | 88.63 | 80.86 |
| TurboParser + conv. rules | - | 80.68 |

- using TSParser alone slightly outperforms our baseline system on French
- Close to the SOTA for DM (89.90 LF on the dev set)

# Will adding more syntactic context help?

**It did for English**

- In our previous work , we showed that adding more syntactic context to a mid-performing transition-based graph parser was highly beneficial for English deep syntax parsing
- Doing so also slightly improved a high performing global model such as TurboSemanticParser's (+0.6 pt).

**So, we used two types of features: *constituent* and *dependency* features:**

- Constituents come from the Berkeley Parser (Petrov et al., 2006).
- Dependencies come from the Mate Parser (Bohnet, 2010) for English and a TAG-based metagrammar, FrMG, (Villemonte De La Clergerie, 2010) for French.

# WILL ADDING MORE SYNTACTIC CONTEXT HELP?

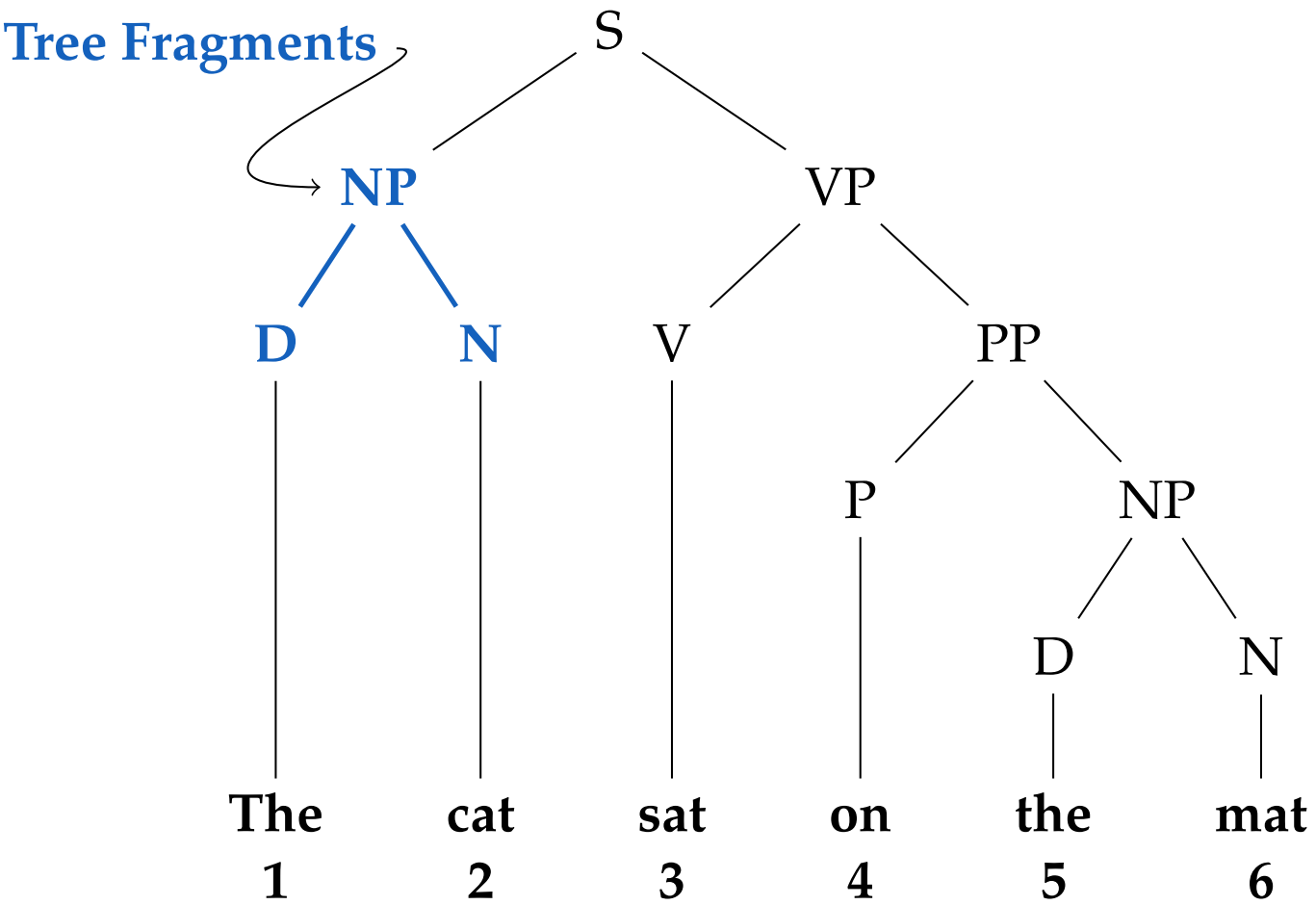
**It did for English**

- In our previous work , we showed that adding more syntactic context to a mid-performing transition-based graph parser was highly beneficial for English deep syntax parsing
- Doing so also slightly improved a high performing global model such as TurboSemanticParser's (+0.6 pt).

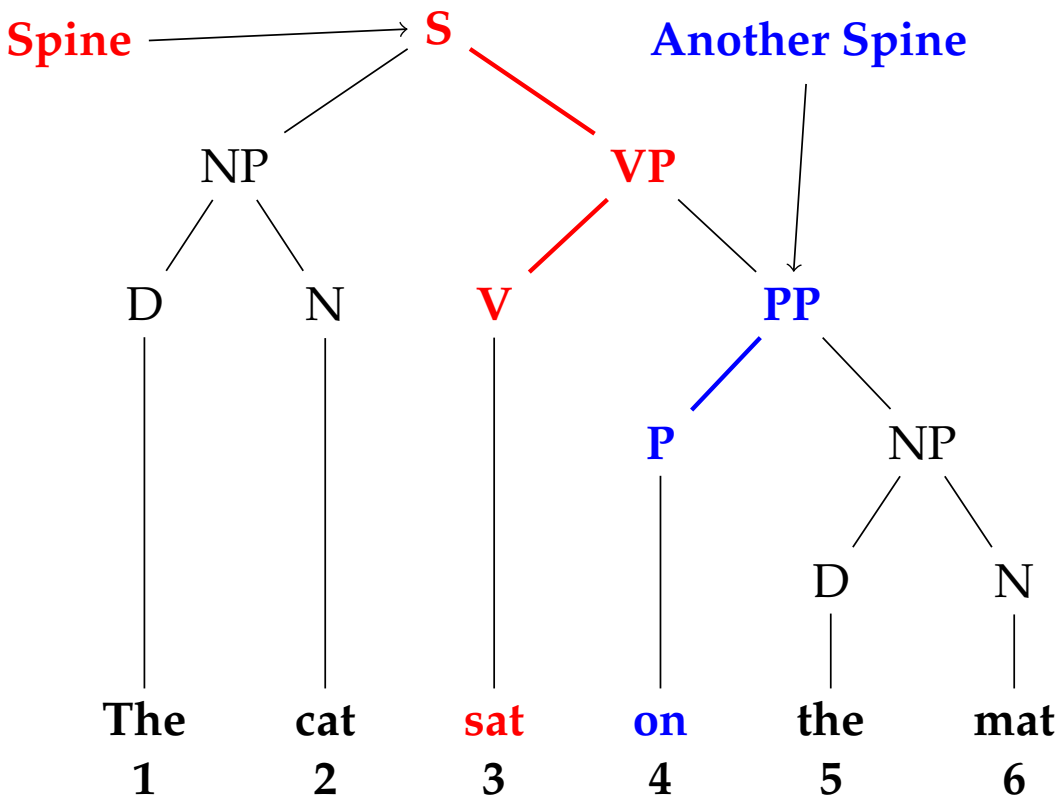
**Correct performance from these parsers on the FTB**

| | BKY | FRMG |
|---|---|---|
| Dev | 80.19 | 83.41 |
| Test | 80.14 | 83.22 |

# Tree Fragments

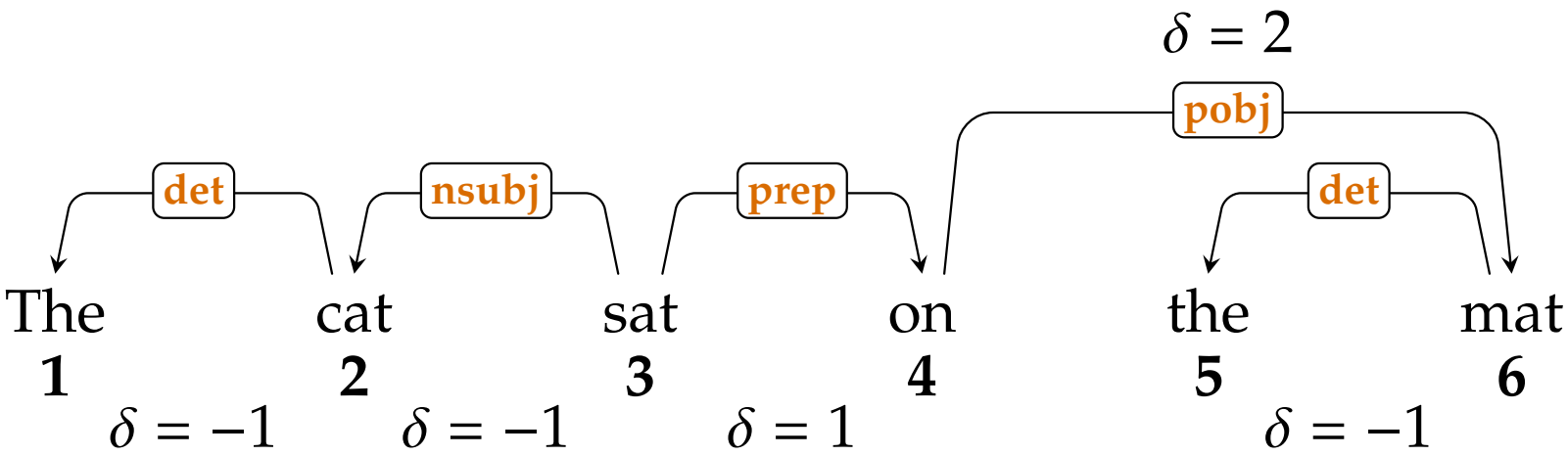


# Spine Fragments



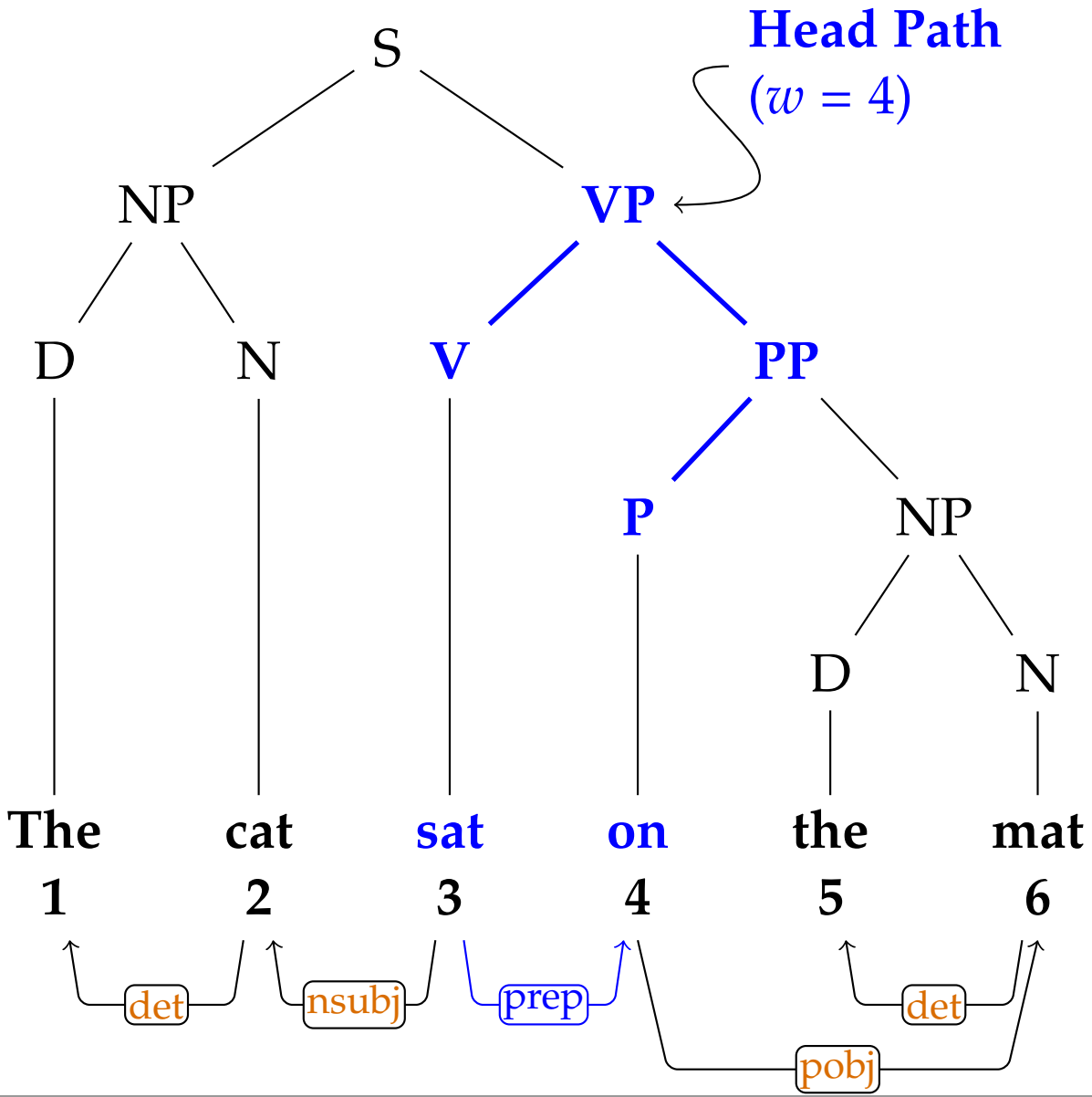

- Path between POS and maximal projection of a head.
- Assigned in a deterministic way (Head-percolation table)

# Dependencies

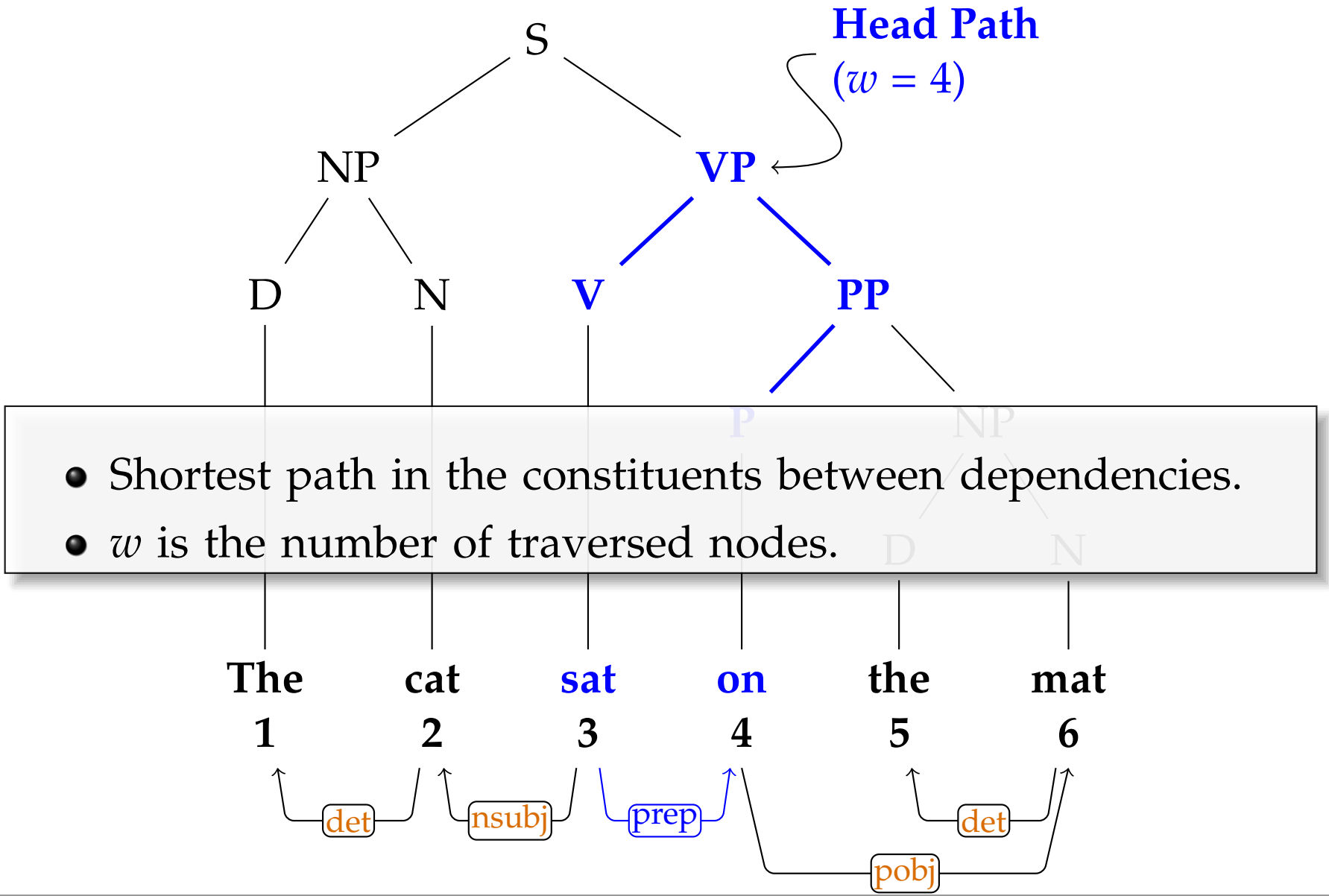



- $\delta$ = directed distance between two words linked by a dependency.
- We use dependency labels.
- We also tested with a pair <head POS, label>.

# Head Paths


Head Path
($w = 4$)
S
NP
VP
D
N
V
PP
P
NP
D
N
The
cat
sat
on
the
mat
1
2
3
4
5
6
det
nsubj
prep
det
pobj

# Head Paths




- Shortest path in the constituents between dependencies.
- $w$ is the number of traversed nodes.

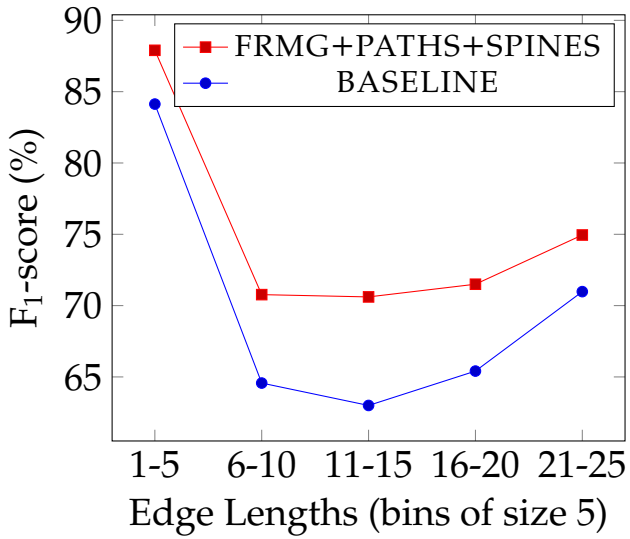
# Impact of syntactic feature

| Dev. set | LP | LR | LF | |
|---|---|---|---|---|
| BASELINE | 83.04 | 78.80 | 80.86 | |
| BKY | 83.63 | 79.67 | 81.60 | +0.74 |
| SPINES | 83.72 | 80.05 | 81.84 | +0.98 |
| PATHS | 84.75 | 81.17 | 82.92 | +2.06 |
| FRMG | 86.50 | 82.74 | 84.58 | **+3.72** |
| FRMG+PATHS+BKY | 86.11 | 83.68 | 84.88 | +4.02 |
| FRMG+PATHS+SPINES | 86.15 | 83.71 | 84.91 | **+4.05** |

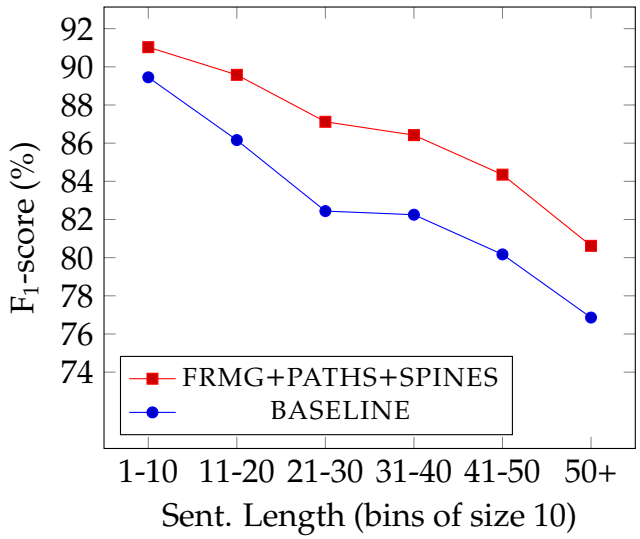
## Highly beneficial

1. Merging topologically different features improves performance
2. FRMG itself is already a kind of mixed model (derivations of elementary trees → dependencies. Extended domain of locality)

# Results Analysis




(a) Long-distance Dep.

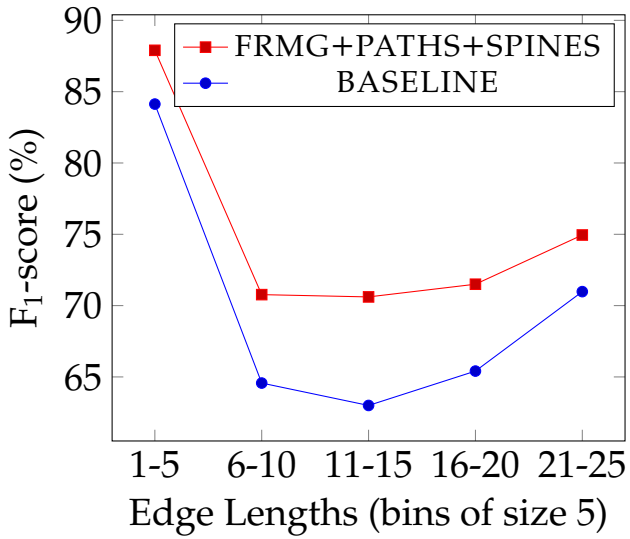



(b) Sentence Lengths

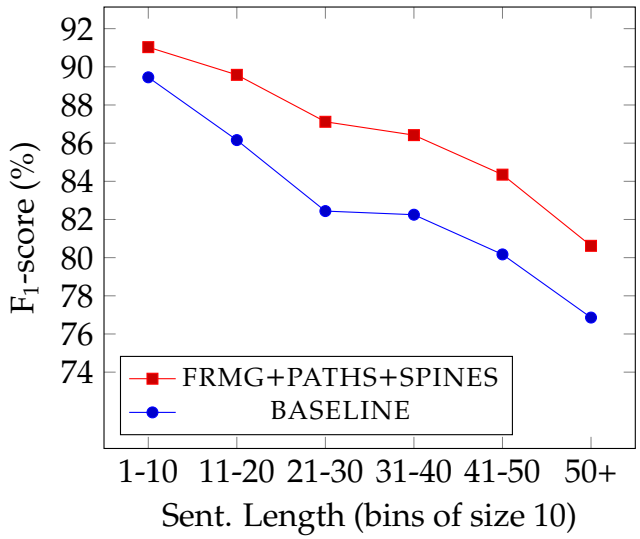
## LDDs: capturing shared subject/object of coordinated verbs

- Argument ellipsis coordination are notoriously difficult to parse
- Providing more syntactic context helps to cope with the lack of coord. structues in the training data (around 15% for DM)
- at least 2 times increase for longer dependencies than shorter ones

# Results Analysis




(c) Long-distance Dep.




(d) Sentence Lengths

## Alleviating the sentence length factor

- small improvement for short sentences 1.5 pt, x 4 for longer ones.
- follows the intuition: providing more context for difficult constructs generalizes over longer sentences (less error propagations, even for a global model)

# Final Results

| test set | DM | DeepFTB |
|---|---|---|
| TSParser+syntactic feat. | **89.70** | **85.38** |
| DSR+ syntactic feat. | 85.66 | 83.38 |
| TSParser baseline | **88.08** | **80.79** |
| DSR, baseline | 83.91 | 76.52 |
| TParser + conversion rules | - | 80.45 |

**For validation, we ran our experiments on a transition-based graph parser with beams and aggressive early updates (**DSR**, (Villemonte De La Clergerie, 2013))**

- Same observed trends apply.
- Because its decision are local, it benefits much more from additional syntactic context

# Final Results

| test set | DM | DeepFTB |
|---|---|---|
| TSParser+syntactic feat. | **89.70** | **85.38** |
| DSR+ syntactic feat. | 85.66 | 83.38 |
| TSParser baseline | **88.08** | **80.79** |
| DSR, baseline | 83.91 | 76.52 |
| TParser + conversion rules | - | 80.45 |

- DM parsers used the same feature set as for the DeepFTB (mate dependencies instead of Frmg's)
- Improvement in both cases
- using topologically-different syntactic features generalizes across languages.

# Final Results

| test set | DM | DeepFTB |
|---|---|---|
| TSParser+syntactic feat. | **89.70** | **85.38** |
| DSR+ syntactic feat. | 85.66 | 83.38 |
| TSParser baseline | **88.08** | **80.79** |
| DSR, baseline | 83.91 | 76.52 |
| TParser + conversion rules | - | 80.45 |

**Deep-syntax parsing of French seems more sensitive to the addition of such features** :

- relatively small data set size?
- Ambitious annotation scheme?
- Still, results are good and encouraging!

# Conclusion

## Regarding the parsing of the Deep-FTB

- We showed it was doable and was performing reasonably well.
- This data set is available under the same conditions as the classic French Treebank (contact Marie Candito)
- Part of groups with same annotation scheme: Deep Sequoia (3k sent), French QuestionBank (2.8k)

## Exciting things are coming up

1. Evolution of the UD scheme toward more semantically oriented graph structures
2. Maybe more multilingual data set with different annotation schemes? which one will the best adapted to the task?
3. In all cases, we're looking forward to this Deep-syntax "revival"!

# Semantic Parsing of French is Near

## Application to French Framenet semantic parsing

- Joint on-going work with LIF (Alexis Nasr, Olivier Michalon) and Marie Candito (ANR Asfalda)
- Goal: automatic prediction of Framenet's frames and roles
- hypothesis: Information that matter to predict roles are of syntactic natures (linking regularities)
- first results: positive impact of deep vs surface syntax
- static evaluation: deep syntactic paths are more regular
- dynamic evaluation: improves syntactic roles prediction

Merci !

(contact: **djame.seddah@paris-sorbonne.fr**)