



# Neural Coreference Resolution with Limited Lexical Context and Explicit Mention Detection for Oral French

---

Loïc Grobol

Université Sorbonne Nouvelle / Inria

CRAC 2019

Minneapolis, 2019-06-07

## Coreference detection for French

---

## Coreference in French

---

Not fundamentally different from other Romance languages like Italian, Spanish...

→ Somewhat different from English because of marked grammatical genders

For now, only a single corpus of a significant size with coreference annotations

- ANCOR (Muzerelle et al. 2014), a corpus of transcribed **spontaneous oral**

An equivalent corpus of *written* French, DEMOCRAT (Landragin 2016), should be released in the coming days

**Disclaimer** In this talk, *mention* includes singleton mentions

(Muzerelle et al. 2014)

- Transcribed spontaneous oral with varying interactivity
- 418 k words, 116 k mentions and 45 k coreference relations

(Muzerelle et al. 2014)

- Transcribed spontaneous oral with varying interactivity
- 418 k words, 116 k mentions and 45 k coreference relations
- “[la toute première] est bien simple [je] voudrais savoir depuis combien [de temps] [vous] habitez à [Orléans]”

(Muzerelle et al. 2014)

- Transcribed **spontaneous** oral with varying interactivity
- 418 k words, 116 k mentions and 45 k coreference relations
- “[la toute première] est bien simple [je] voudrais savoir depuis combien [de temps] [vous] habitez à [Orléans]”
- “bah [j’] aime euh [j’] aime bien les disons les euh le ciné- [le cinéma actuel] euh [le cinéma] euh euh un petit symbolique”

(Muzerelle et al. 2014)

- Transcribed **spontaneous** oral with varying interactivity
- 418 k words, 116 k mentions and 45 k coreference relations
- “[la toute première] est bien simple [je] voudrais savoir depuis combien [de temps] [vous] habitez à [Orléans]”
- “bah [j’] aime euh [j’] aime bien les disons les euh le ciné- [le cinéma actuel] euh [le cinéma] euh euh un petit symbolique”
- “bah il y a il y a des il y a il y a des il y a [des termes]”

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries



alors les bourses oui je sais un petit peu comment ça se passe et parce que moi je m'intéresse un peu aux questions financières euh j'ai appris récemment parce qu'au rectorat j'avais fait la connaissance du chef du service des bourses alors euh il y a eu un changement de système entre l'année dernière et cette année auparavant les il y avait des un comité de bourses ou je ne sais quoi qui se réunissait une commission qui décidait bon ou le dans le service des bourses on voyait un dossier était déposé on voyait un tel bon euh aura une bourse on décidait non pas au pif mais en tenant compte de certains correctifs qui sont euh vous savez qu'en France c'est lié aux impôts tout simplement si bien on vous a alors les impôts euh il y a deux deux catégories de Français rendus compte dans le système des impôts il y a le non-salarié et les salariés alors les non-salariés eux ont touchent donc leur traitement qui est déclaré intégralement à la direction des impôts donc sur les déclarations d'impôts on ne peut pas retirer un centime et on se retrouve à la fin de la colonne avec un revenu imposable qui est effectivement celui que vous devez déclarer les catégories qui non-salariées les commerçants les artisans les professions libérales eux peuvent au contraire on peut pas les contrôler les agriculteurs et ils mettent non pas ce qu'ils veulent faut pas exagérer mais enfin tout le monde sait et il y a des correctifs pour ça d'ailleurs tout le monde sait que euh s'il y a un agriculteur déclare tant ben c'est qu'il a gagné deux fois plus trois fois plus ou un commerçant et cetera et on en parle beaucoup actuellement en France hein des des réformes des impôts tout ça c'est très dé- c'est un problème difficile alors les bourses sont accordées suivant le revenu imposable donc voyez que

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words
- Disfluencies

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words
- Disfluencies
- Transcription noise

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words
- Disfluencies
- Transcription noise
  - c' est qu' il a gagné deux foisplus trois fois plus



## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words
- Disfluencies
- Transcription noise
  - c' est qu' il a gagné deux foisplus trois fois plus
  - ben e(lle) elle est

## Spontaneous oral

---

- Overly long utterances with no clear sentence boundaries
- Lots of short non-sentence utterances
  - Over 2000 occurrences of “hm hm”
- Document size discrepancies
  - From 4 utterances/9 words to 1200 utterances/16 000 words
- Disfluencies
- Transcription noise
  - c' est qu' il a gagné deux foisplus trois fois plus
  - ben e(lle) elle est

Some of these are easy to clean, other really hard, but in any case, this is very different from the type of documents found in other corpora and makes off-the-shelf preprocessing tools a lot less reliable than usual

# End-to-end neural coreference detection

---

## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**

## Span-based antecedent finding

---

Sam Vimes sighed when he heard the scream

## Span-based antecedent finding

---

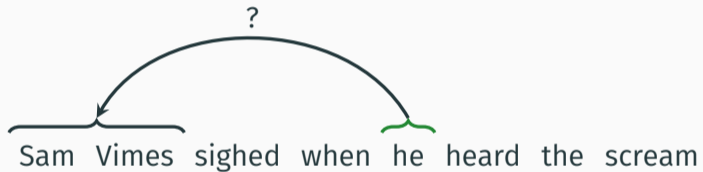
Sam Vimes sighed when  he heard the scream

## Span-based antecedent finding

---

 Sam Vimes sighed when  he heard the scream

## Span-based antecedent finding





## Span-based antecedent finding

---

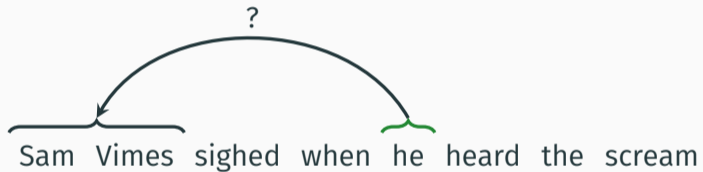
Sam Vimes sighed when  he heard the scream

## Span-based antecedent finding

---

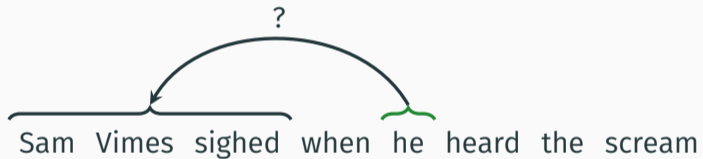


## Span-based antecedent finding



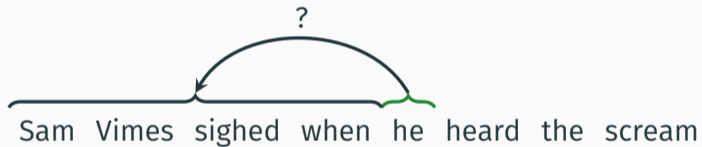
## Span-based antecedent finding

---



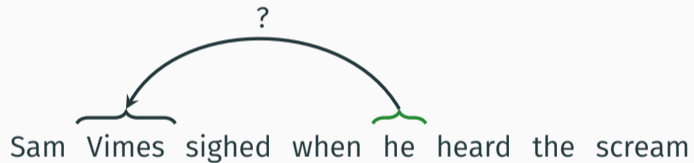
## Span-based antecedent finding

---



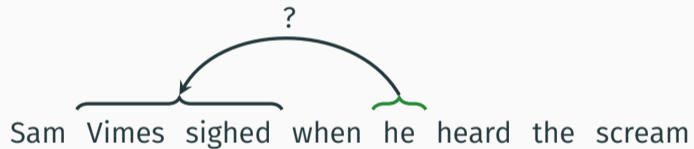
## Span-based antecedent finding

---



## Span-based antecedent finding


---



## Span-based antecedent finding

---

Sam Vimes sighed when he heard the scream




The diagram illustrates the process of finding the antecedent for the pronoun 'he' in the sentence 'Sam Vimes sighed when he heard the scream'. A horizontal line spans from the start of 'S' to the end of 'e' in 'he'. A curved arrow points from the end of 'he' back to the start of 'sighed'. A question mark is positioned above the arrow, indicating the task of identifying the antecedent. A green underline is present under the word 'he'.



## Span-based antecedent finding

---

Sam Vimes sighed when he heard the scream



## Span-based antecedent finding

---


Sam Vimes sighed when he heard the scream



## Span-based antecedent finding

---

Sam Vimes sighed when he heard the scream



## Span-based antecedent finding

---



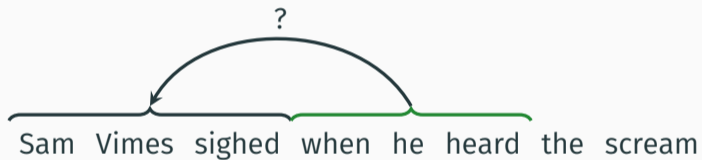
## Span-based antecedent finding

---



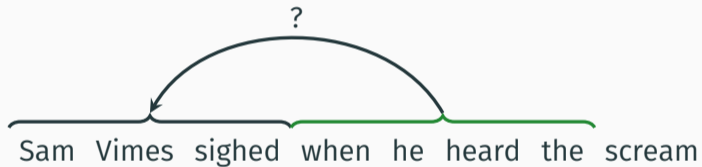
## Span-based antecedent finding

---



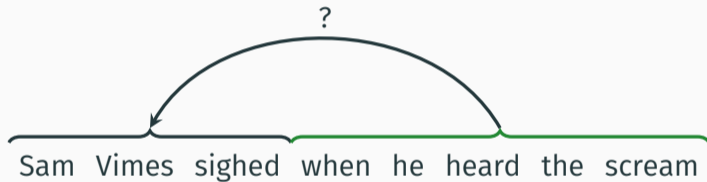
## Span-based antecedent finding

---



## Span-based antecedent finding

---





## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**

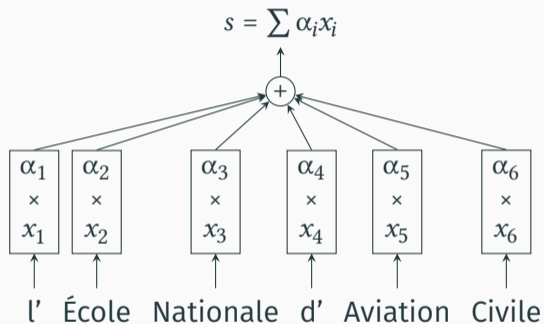
## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

## Soft-head attention



Soft-head

Attention weights  $\alpha = \text{softmax}(\text{FFNN}(\mathbf{x}))$

Word-level representations

## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

## End-to-end coreference resolution

---

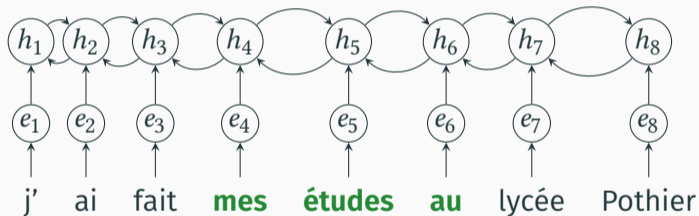
Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

In details

- Words are represented at the sentence level by the hidden states of a BiLSTM

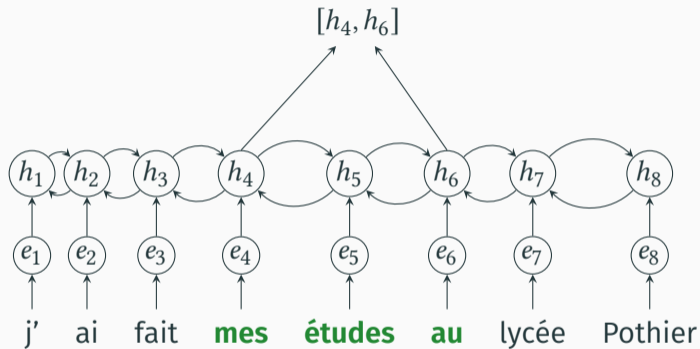
## Encoding text spans



Recurrent representations

Word-level representations

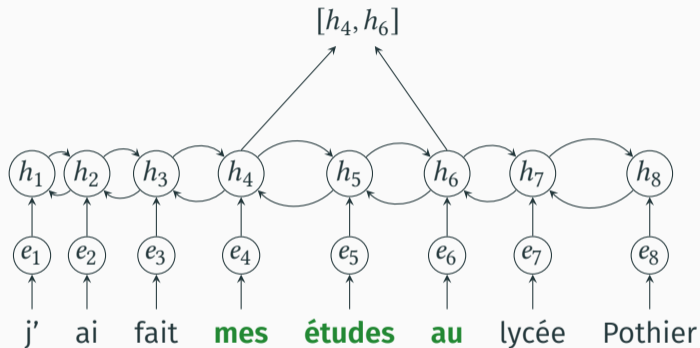
## Encoding text spans



Recurrent representations

Word-level representations

## Encoding text spans



Recurrent representations

Word-level representations

Final span representation:  $g_{4,6} = [h_4, s, h_6, \phi]$



## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

In details

- Words are represented at the sentence level by the hidden states of a BiLSTM

## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

In details

- Words are represented at the sentence level by the hidden states of a BiLSTM
- The total score for a pair is the sum of the **mention scores** of both spans and the **antecedent score** of the pair:  $a(i, j) = s_m(g_i) + s_m(g_j) + s_a(g_i, g_j, \psi)$

## End-to-end coreference resolution

---

Lee, He, Lewis, et al. (2017) ditch the preprocessing (aka syntactic parsing and rule-based mention detection)

- Antecedent scoring on **spans**
- Use **attention** as a substitute for the mention head features of previous works

In details

- Words are represented at the sentence level by the hidden states of a BiLSTM
- The total score for a pair is the sum of the **mention scores** of both spans and the **antecedent score** of the pair:  $a(i, j) = s_m(g_i) + s_m(g_j) + s_a(g_i, g_j, \psi)$

Lee, He, and Zettlemoyer (2018) add refining steps to take the antecedent context into account in spans representations

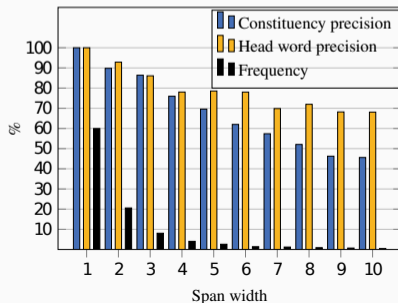
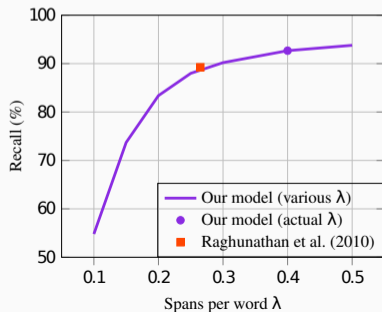
## Successes

State-of-the-art on CoNLL-2012 for English.

→ Improves the previous SotA by a margin of 1.5 points (more with refinements)

Successfully remove the reliance on syntactic parsing

- Detects more mentions than rule-based detectors using predicted trees
- Soft-head corresponds to predicted syntactic head *where it is relevant*



## Problems

---

End-to-end processing is computationally very heavy

- Aggressive pruning is performed at every step

End-to-end processing is computationally very heavy

- Aggressive pruning is performed at every step
  - Truncate documents, prune spans, limit the number antecedent candidates

End-to-end processing is computationally very heavy

- Aggressive pruning is performed at every step
  - Truncate documents, prune spans, limit the number antecedent candidates
  - Harder to implement and train

End-to-end processing is computationally very heavy

- Aggressive pruning is performed at every step
  - Truncate documents, prune spans, limit the number antecedent candidates
  - Harder to implement and train
- Still quite demanding, especially for documents with overly long utterances



## Problems

---

End-to-end processing is computationally very heavy

- Aggressive pruning is performed at every step
  - Truncate documents, prune spans, limit the number antecedent candidates
  - Harder to implement and train
- Still quite demanding, especially for documents with overly long utterances

Does not perform mention detection, but only **antecedent** detection :  
non-mentions and first mentions are treated the same way

→ Doesn't allow working with singleton mentions

## Adapting E2EC

---

Explicit mention detection

## Adaptations

---

Explicit mention detection

- Train (and evaluate) explicitly on mention detection

## Explicit mention detection

---

Sam Vimes sighed when he heard the scream


## Explicit mention detection

---

 Sam Vimes sighed when he heard the scream

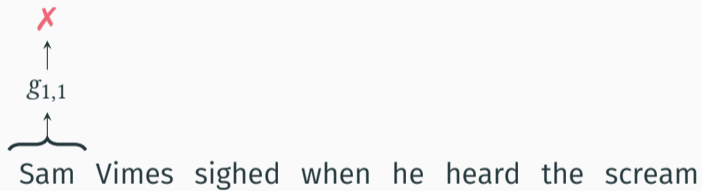
## Explicit mention detection

---

$g_{1,1}$   
  
Sam Vimes sighed when he heard the scream

## Explicit mention detection

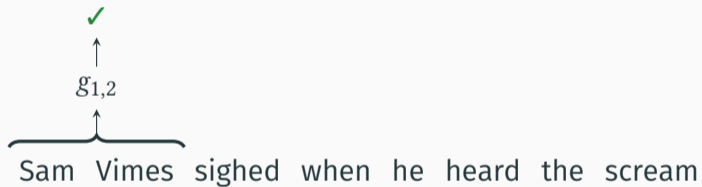
---





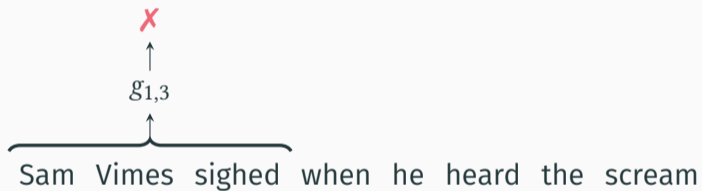
## Explicit mention detection

---



## Explicit mention detection

---



## Explicit mention detection

---

A diagram illustrating explicit mention detection. The sentence "Sam Vimes sighed when he heard the scream" is shown. A bracket underlines "Sam Vimes". An arrow points from the bracket to the label  $g_{2,2}$ . Another arrow points from  $g_{2,2}$  to a red 'X' above the word "sighed".

Sam Vimes sighed when he heard the scream

## Explicit mention detection

---

Sam Vimes sighed when he heard the scream



# Adaptations

---

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations
  - Improves pair modelling for antecedent finding (Lassalle and Denis 2013)



## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations
  - Improves pair modelling for antecedent finding (Lassalle and Denis 2013)

## Use a fixed context instead of sentences

- Instead of deriving word representations from whole utterances, use the immediate context

alors les bourses oui je sais un petit peu comment ça se passe et parce que moi je m'intéresse un peu aux questions financières euh j'ai appris récemment parce qu'au rectorat j'avais fait la connaissance du chef du service des bourses alors euh il y a eu un changement de système entre l'année dernière et cette année auparavant les il y avait des un comité de bourses ou je ne sais quoi qui se réunissait une commission qui décidait bon ou le dans le service des bourses on voyait un dossier était déposé on voyait un tel bon euh aura une bourse on décidait non pas au pif mais en tenant compte de **certaines correctifs** qui sont euh vous savez qu'en France c'est lié aux impôts tout simplement si bien on vous a alors les impôts euh il y a deux deux catégories de Français rendus compte dans le système des impôts il y a le non-salarié et les salariés alors les non-salariés eux ont touchent donc leur traitement qui est déclaré intégralement à la direction des impôts donc sur les déclarations d'impôts on ne peut pas retirer un centime et on se retrouve à la fin de la colonne avec un revenu imposable qui est effectivement celui que vous devez déclarer les catégories qui non-salariées les commerçants les artisans les professions libérales eux peuvent au contraire on peut pas les contrôler les agriculteurs et ils mettent non pas ce qu'ils veulent faut pas exagérer mais enfin tout le monde sait et il y a des correctifs pour ça d'ailleurs tout le monde sait que euh s'il y a un agriculteur déclare tant ben c'est qu'il a gagné deux fois plus trois fois plus ou un commerçant et cetera et on en parle beaucoup actuellement en France hein des des réformes des impôts tout ça c'est très dé- c'est un problème difficile alors les bourses sont accordées suivant le revenu imposable donc voyez que

alors les bourses oui je sais un petit peu comment ça se passe et parce que moi je m'intéresse un peu aux questions financières euh j'ai appris récemment parce qu'au rectorat j'avais fait la connaissance du chef du service des bourses alors euh il y a eu un changement de système entre l'année dernière et cette année auparavant les il y avait des un comité de bourses ou je ne sais quoi qui se réunissait une commission qui décidait bon ou le dans le service des bourses on voyait un dossier était déposé on voyait un tel bon euh aura une bourse on décidait non pas au pif mais en tenant compte de **certains correctifs** qui sont euh vous savez qu'en France c'est lié aux impôts tout simplement si bien on vous a alors les impôts euh il y a deux deux catégories de Français rendus compte dans le système des impôts il y a le non-salarié et les salariés alors les non-salariés eux ont touchent donc leur traitement qui est déclaré intégralement à la direction des impôts donc sur les déclarations d'impôts on ne peut pas retirer un centime et on se retrouve à la fin de la colonne avec un revenu imposable qui est effectivement celui que vous devez déclarer les catégories qui non-salariées les commerçants les artisans les professions libérales eux peuvent au contraire on peut pas les contrôler les agriculteurs et ils mettent non pas ce qu'ils veulent faut pas exagérer mais enfin tout le monde sait et il y a des correctifs pour ça d'ailleurs tout le monde sait que euh s'il y a un agriculteur déclare tant ben c'est qu'il a gagné deux fois plus trois fois plus ou un commerçant et cetera et on en parle beaucoup actuellement en France hein des des réformes des impôts tout ça c'est très dé- c'est un problème difficile alors les bourses sont accordées suivant le revenu imposable donc voyez que

aura une bourse on décidait non pas au pif mais en tenant compte de certains correctifs qui sont euh vous savez qu' en France c' est lié aux impôts tout simplement

# Adaptations

---

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations
  - Improves pair modelling for antecedent finding (Lassalle and Denis 2013)

## Use a fixed context instead of sentences

- Instead of deriving word representations from whole utterances, use the immediate context

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations
  - Improves pair modelling for antecedent finding (Lassalle and Denis 2013)

## Use a fixed context instead of sentences

- Instead of deriving word representations from whole utterances, use the immediate context
- Attach context embeddings to spans representations to help get through disfluencies

# Adaptations

---

## Explicit mention detection

- Train (and evaluate) explicitly on mention detection
- Learn the **type** of mentions (PR vs NP)
  - Improve span representations
  - Improves pair modelling for antecedent finding (Lassalle and Denis 2013)

## Use a fixed context instead of sentences

- Instead of deriving word representations from whole utterances, use the immediate context
- Attach context embeddings to spans representations to help get through disfluencies

And some tuning of the specifics

# Experiments

---



## Sequentially training

- First train on mention detection, then on antecedent scoring
- The span representations are shared between the two tasks and trained continuously
- The hope is that learning the easiest task will help learning the more complicated one.

See the paper for the gory details

### Mention detection

<b>System</b>	<b>P</b>	<b>R</b>	<b>F</b>
Grobel et al. (2017)	57.28	77.07	65.72
Godbert and Favre (2017)	90.05	87.86	88.94
Our model	82.99	89.07	85.87

Error analysis suggests that a lot of the false positive are boundary problems

## Coreference resolution

System	MUC			B <sup>3</sup>			CEAF <sub>e</sub>			CoNLL	BLANC		
	P	R	F	P	R	F	P	R	F	Avg.	P	R	F
(1)	—	—	63.5	—	—	83.8	—	—	79.0	75.3	—	—	67.4
(2)	—	—	—	—	—	—	—	—	—	—	—	—	65.7
Our model	72.3	47.7	57.3	89.7	71.0	79.2	72.8	86.0	79.4	72.0	78.2	60.1	65.7

1. Désoyer et al. (2015)
2. Godbert and Favre (2017)

## Perspectives

---

## System improvements

---

Go *back* to more **explicit** processing of documents

- Oral preprocessing: denoising, sentence (period) segmentation...
- Mention detection post-processing: remove mutually contradictory mentions
- Explicit mention-new detection

Use **features** from other systems

- Named entity recognition
- Syntactic parsing

Explore other training modalities

- Quasi-simultaneous multi-task learning
- Curriculum training

Ongoing work in testing the configuration on other corpora

- OntoNotes
  - Replacing the mention detection task by noun phrase detection
- Comparable corpora in other languages
  - AnCora (Taulé et al. 2008)
  - ARRAU (Poesio and Artstein 2008)
  - PCC (Ogrodniczuk et al. 2016)
  - GUM (Zeldes 2017)
  - ...
- DEMOCRAT (Landragin 2016)

And run comprehensive ablation studies to see if the resources demands can be further lowered.

A prototype for coreference detection in French

- Based on the state of the art neural end-to-end system
- Adapted to take into account a more general paradigm : long documents, noisy input, singleton mentions...
- A baseline for coreference detection for (spontaneous oral) French

A usable interface and pretrained models coming soon.

Get in touch at [loic.grobo1@gmail.com](mailto:loic.grobo1@gmail.com) !

## **Appendices**



# References i

Désoyer, Adèle, Frédéric Landragin, Isabelle Tellier, Anaïs Lefeuvre, and Jean-Yves Antoine (May 2015).

“Coreference Resolution for Oral Corpus: a machine learning experiment with ANCOR corpus”. In: *Traitement Automatique des Langues* 55.2.

URL: <https://halshs.archives-ouvertes.fr/halshs-01153297>.

Godbert, Elisabeth and Benoît Favre (June 2017). “Détection de coréférences de bout en bout en français”.

In: *Actes de la 24e Conférence sur le Traitement Automatique des Langues Naturelles*. TALN 2017.

URL: <https://hal.archives-ouvertes.fr/hal-01687116>.

GroboL, Loïc, Isabelle Tellier, Éric De La Clergerie, Marco Dinarelli, and Frédéric Landragin (June 2017).

“Apports des analyses syntaxiques pour la détection automatique de mentions dans un corpus de français oral”.

In: *Actes de la 24e Conférence sur le Traitement Automatique des Langues Naturelles*. TALN 2017. URL: <https://hal.inria.fr/hal-01558711>.

Landragin, Frédéric (2016). “Description, Modélisation et Détection Automatique Des Chaînes de Référence (DEMOCRAT)”. In: *Bulletin de l'AFIA* 92.

Lassalle, Emmanuel and Pascal Denis (Aug. 2013). “Improving Pairwise Coreference Models through Feature Space Hierarchy Learning”. In:

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

URL: <https://www.aclweb.org/anthology/papers/P/P13/P13-1049/>.

Lee, Kenton, Luheng He, Mike Lewis, and Luke Zettlemoyer (Sept. 2017). “End-to-End Neural Coreference Resolution”.

In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2017.

URL: <https://www.aclweb.org/anthology/D17-1018>.

Lee, Kenton, Luheng He, and Luke Zettlemoyer (2018). “Higher-Order Coreference Resolution with Coarse-to-Fine Inference”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

NAACL HLT 2018 (New Orleans, Louisiana). Vol. 2. URL: <http://aclweb.org/anthology/N18-2108>.

URL: <http://aclweb.org/anthology/N18-2108>.

Muzerelle, Judith et al. (2014). “ANCOR Centre, a Large Free Spoken French Coreference Corpus: Description of the Resource and Reliability Measures”.

In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. LREC 2014.

URL: <https://hal.archives-ouvertes.fr/hal-01075679>.

## References ii

- Ogrodniczuk, Maciej, Katarzyna Głowińska, Mateusz Kopeć, Agata Savary, and Magdalena Zawistawska (2016). "Polish Coreference Corpus".  
In: *Human Language Technology. Challenges for Computer Science and Linguistics*.
- Poesio, Massimo and Ron Artstein (2008). "Anaphoric Annotation in the ARRAU Corpus".  
In: *Proceedings of the International Conference on Language Resources and Evaluation*.  
URL: <http://ict.usc.edu/pubs/Anaphoric%20Annotation%20in%20the%20ARRAU%20Corpus.pdf>.
- Taulé, Mariona, M. Antònia Martí, and Marta Recasens (May 2008). "AnCora: Multilevel Annotated Corpora for Catalan and Spanish".  
In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-08)*.  
URL: [http://www.lrec-conf.org/proceedings/lrec2008/pdf/35\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/35_paper.pdf).
- Zeldes, Amir (Sept. 1, 2017). "The GUM Corpus: Creating Multilayer Resources in the Classroom". In: *Language Resources and Evaluation* 51.3.  
URL: <https://doi.org/10.1007/s10579-016-9343-x>.

## Acknowledgements

This work is part of the “Investissements d’Avenir” overseen by the French National Research Agency ANR-10-LABX-0083 (Labex EFL).

This work has been supported by the ANR DEMOCRAT (Description et modélisation des chaînes de référence: outils pour l’annotation de corpus et le traitement automatique) project ANR-15-CE38-0008.



*@democrat*

## License



This document is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0)  
([creativecommons.org/licenses/by/4.0](https://creativecommons.org/licenses/by/4.0))

© 2019, Loïc Grobol <[loic.grobol@gmail.com](mailto:loic.grobol@gmail.com)>  
[lattice.cnrs.fr/Grobol-Loic](https://lattice.cnrs.fr/Grobol-Loic)