

Improving coreference resolution with automatically predicted prosodic information

Ina Rösiger, Sabrina Stehwien
Arndt Riester, Thang Vu

University of Stuttgart
Institute for Natural Language Processing (IMS)

September 07, 2017



University of Stuttgart
Germany



Institut für
Maschinelle
Sprachverarbeitung

Coreference resolution

Grouping references to the same discourse entities together

President Clinton has signed into law a bill allowing US exports of food and medicine to Cuba . Nevertheless, Mr. Clinton says he is not satisfied with the measure . The new law bars the US government and US banks from providing funds for such exports at the insistence of Cuba's congressional critics. It also prevents Mr. Clinton or his successor from easing restrictions on travel to the Communist country .

Coreference resolution

- Highly active NLP area
- Task: partition NPs in a document into coreference chains
- Different approaches: most are statistical
- Text-based features:
part-of-speech, syntactic parses, morphological information
- Systems trained on written text do not perform as well on spoken text

Why use prosody for coreference resolution?

John has an old cottage. Last year he reconstructed the shed.

← *coreferent?*

Why prosodic prominence matters

John has an old cottage.
Last year he reconstructed the
SHED.



cottage



part of shed
←

⇒ the cottage and the shed do not corefer

Why prosodic prominence matters

John has an old cottage.
Last year he **reconSTRUCTed**
the shed.



cottage = shed

⇒ the cottage and the shed corefer

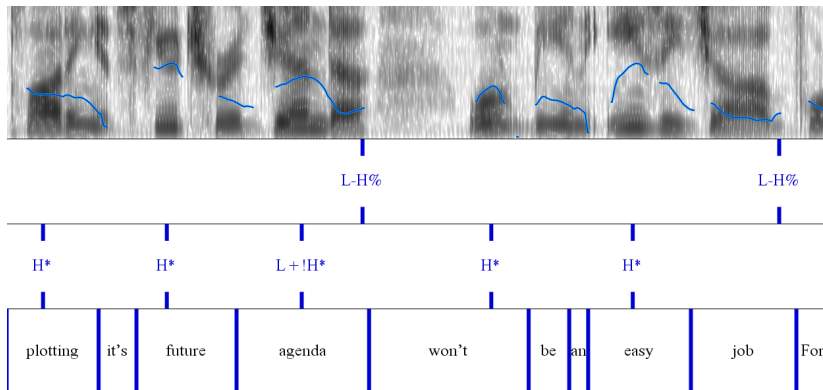
Motivation

- Prosody can give clues where transcript is ambiguous
 - Accentuation can distinguish given and new information
 - Pilot study for German [Rösiger and Riester 2015](#)
 - shown that prosodic information can help coreference resolution
 - based on manually annotated pitch accents and boundary tones
 - added prosodic information to a set of text-based, predicted features
 - Practical applications would rely on automatically predicted prosodic information
- focus of this work

Prosodic features for coreference resolution

- We use pitch accents and phrase boundaries
- Phrase boundaries are used to derive the nuclear accent
 - last accent in intonation phrase
 - perceived as most prominent
- Two binary features used in the resolver:
 - pitch accent presence
 - nuclear accent presence

Prosodic events: ToBI example



Accent type and NP length

- **Pitch accents** are helpful clues for **short NPs**
 - make it more likely for the NP to contain new information
 - *the* **SHED**, *President* **CLINTON**, ...
- **Nuclear accents** are helpful for **long NPs**
 - they almost always have at least one pitch accent
 - *a* **BILL** *allowing* **US EXPORTS** *of food and medicine to* **CUBA**

Data

- DIRNDL anaphora corpus [Eckart et al. 2012](#), [Björkelund et al. 2014](#)
- consists of 4.5 hours of German radio news
- 13 male and 7 female speakers
- manually annotated for coreference and prosodic events
- we use the official training, dev and test set splitting

Coreference resolver

- Data-driven coreference resolver:
 - IMS HotCoref DE [Rösiger and Kuhn 2016](#)
 - state-of-the-art resolver for German
 - structured perceptron that models coreference in a document as a directed rooted tree, following [Björkelund and Kuhn 2014](#)
 - standard features: string-matching, part-of-speech, constituent trees, morphological information, etc.
- Performance is evaluated with the CoNLL score

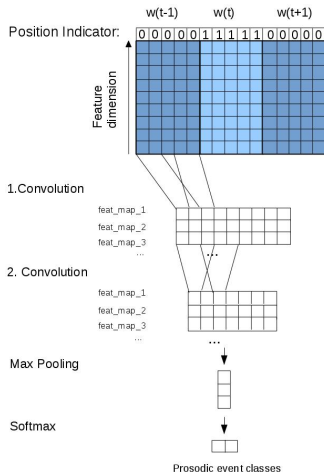
Goal: **completely automatic preprocessing**

All features for the coreference resolver were obtained using automatic NLP methods

CNN-based prosodic event detection

Stehwien and Vu 2017

- supervised learning task: each word is labelled as carrying a prosodic event or not
- feature matrix: frame-based representation of audio signal
- 2 convolution layers
- max pooling finds most salient features
- resulting feature maps concatenated to one feature vector
- softmax layer: 2 units for binary classification



Method

1. Automatic extraction of text-based features
2. Prosodic event detector is applied to the DIRNDL corpus to obtain pitch accents and phrase boundaries (separately)
 - Model pre-trained on Boston University Radio News Corpus
[Ostendorf et al. 1995](#)
 - Prediction accuracy on DIRNDL:
 - Pitch accents: 81.9%
 - Phrase boundaries: 85.5%
3. Coreference resolver is trained using the training and development split of DIRNDL
4. Performance is evaluated on the DIRNDL test set

Experimental setup

- Three settings: coreference resolver ...
 - (a) ... trained and tested using manual prosodic labels (short *gold*),
 - (b) ... trained on manual prosodic information, but tested on automatic labels (short *gold/auto*) and
 - (c) ... trained and tested using automatically predicted prosodic labels (short *auto*).
- Two versions:
 - *short NPs*: feature only for NPs of length 3 or less
 - *all NPs*: feature used on all NPs⇒ evaluation always on all NPs

Results

Pitch accent presence:

Baseline	46.11	
+ Accent	short NPs	all NPs
+ Pitch accent presence gold	53.99	49.68
+ Pitch accent presence gold/auto	52.63	50.08
+ Pitch accent presence auto	49.13	49.01

Nuclear accent presence:

Baseline	46.11	
+ Accent	short NPs	all NPs
+ Nuclear accent presence gold	48.63	52.12
+ Nuclear accent presence gold/auto	48.46	51.45
+ Nuclear accent presence auto	48.01	50.64

- significant improvement in all settings¹
- performance of the three settings: *gold* > *gold/auto* > *auto*

¹Wilcoxon signed rank test, $p < 0.01$

Effect of pitch accent and nuclear accent presence

- Pitch accent presence:
 - for long NPs is not helpful: almost always accented
 - including them (*all NP*) limits the feature's informativity
 - on short NPs, a pitch accent makes it more likely for the NP to contain new information
 - best score in *short NP* setting
 - best experimental result (ratio short:long NPs = 3:1)
- Nuclear accent presence:
 - only a few short NPs have a nuclear accent
 - feature is less helpful in the *short NP* setting
 - more meaningful for long NPs
 - best score in *all NP* setting

DIRNDL example

EXPERTEN der Großen **KOALITION** haben sich auf ein Niedriglohn-Konzept **VERSTÄNDIGT**. Die strittigen Themen sollten bei der nächsten Spitzenrunde **der Koalition** **ANGESPROCHEN** werden.

*EN: Experts within the **the grand coalition** have agreed on a strategy to address [problems associated with] low income. At the next meeting, **the coalition** will talk about the controversial issues.*

Conclusion and future work

- Observations of pilot study confirmed
- Prosodic information has a positive effect even when predicted by a system
(despite lower quality of the prosodic annotations)
- Future work:
 - include the available lexicosyntactic information for automatic prosodic labelling
 - fully automatic system based on ASR output

Thank you!



ina.roesiger
sabrina.stehwien
arndt.riester
thang.vu

@ims.uni-stuttgart.de

References



Ina Rösiger and Arndt Riester (2015)

Using prosodic annotations to improve coreference resolution of spoken text
Proceedings of ACL-IJCNLP



Ina Rösiger and Jonas Kuhn (2016)

IMS HotCoref DE: A data-driven co-reference resolver for German
Proceedings of LREC



Sabrina Stehwien and Ngoc Thang Vu (2017)

Prosodic event recognition using convolutional neural networks with context information
Proceedings of Interspeech



Stefan Baumann and Arndt Riester (2013)

Coreference, lexical givenness and prosody in German
Lingua



Anders Björkelund and Jonas Kuhn (2014)

Learning structured perceptrons for coreference resolution with latent antecedents and non-local features
Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics



Anders Björkelund, Kerstin Eckart, Arndt Riester, Nadja Schauffler, Katrin Schweitzer (2014)

The extended DIRNDL corpus as a resource for automatic coreference and bridging resolution
Proceedings of LREC



Mari Ostendorf, Patti Price, Stefanie Shattuck-Hufnagel (1995)

The Boston University Radio News Corpus
Technical Report ECS-95-001, Boston University