Modeling the lifespan of entities with application to coreference resolution

Marie-Catherine de Marneffe
The Ohio State University
Linguistics

Marta Recasens

Christopher Potts
Stanford Linguistics
Discourse referents

[a system designed to keep track of] all the individuals, that is, events, objects, etc., mentioned in the text and, for each individual, record whatever is said about it

Discourse referents
Karttunen (1976)
Anaphoric relations

Joan thanked Sue for all the help she had given.
Anaphoric relations

Joan thanked Sue for all the help she had given.

? Joan
? Sue
Anaphoric relations

Joan thanked Sue for all the help she had received.
Anaphoric relations

Joan critiqued Sue for all the help she had received.
Pronoun resolution

[ Hobbs 1976-1978 ]

Titus: What villain boy! Barrst me my way in Rome?

Mutius: Help, Lucius, help.

[ He kills him. ]

Who kills whom? Lucius kills Mutius.
Pronoun resolution

[Hobbs 1976-1978]

Titus: What villain boy! Barrst me my way in Rome?

Mutius: Help, Lucius, help.

[He kills him.]

Who kills whom? Lucius kills Mutius. Titus
Mentions refer to discourse entities

<table>
<thead>
<tr>
<th>Discourse entities</th>
<th>Titus</th>
<th>Mutius</th>
<th>Lucius</th>
<th>Rome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentions</td>
<td>Titus</td>
<td>Mutius</td>
<td>Lucius</td>
<td>Rome</td>
</tr>
<tr>
<td></td>
<td>me</td>
<td>villain boy</td>
<td>Lucius</td>
<td>Rome</td>
</tr>
<tr>
<td></td>
<td>He</td>
<td>him</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Titus: What villain boy! Barrst me my way in Rome?
Mutius: Help, Lucius, help.

[He kills him]
Some mentions co-refer

Titus: What villain boy! Barrst me my way in Rome?

Mutius: Help, Lucius, help.

[He kills him]
This is hard for computers!

[Durrett & Klein 2013]

Performance on CoNLL-2011 test set

<table>
<thead>
<tr>
<th></th>
<th>CoNLL score</th>
</tr>
</thead>
<tbody>
<tr>
<td>STANFORD</td>
<td>56.65</td>
</tr>
<tr>
<td>IMS</td>
<td>58.26</td>
</tr>
<tr>
<td>BERKELEY</td>
<td>60.13</td>
</tr>
</tbody>
</table>
Not so many years ago, when photographers needed to adjust a photo, they’d have to spend time in the darkroom tinkering with equipment and chemicals. Then along came digital imagery and the magic of Photoshop, which brought photo manipulation out of the dark. Nowadays, smartphone apps can manipulate photos in myriad ways.

Photoshop is the granddaddy of photo manipulation on a smartphone. The Photoshop Touch app is a touch-screen version of the original. It’s powerful and can apply many of the effects you have seen in magazines and on the Web. That includes adjustments like the tilt-shift effect, which can make a townscape look like a miniature model, as well as better-known
Not so many years ago, when photographers needed to adjust a photo, they’d have to spend time in the darkroom tinkering with equipment and chemicals. Then along came digital imagery and the magic of Photoshop, which brought photo manipulation out of the dark. Nowadays, smartphone apps can manipulate photos in myriad ways.

Photoshop is the granddaddy of photo manipulation on a smartphone. The Photoshop Touch app is a touch-screen version of the original. It’s powerful and can apply many of the effects you have seen in magazines and on the Web. That includes adjustments like the tilt-shift effect, which can make a townscape look like a miniature model, as well as better-known
Singleton entities

Not so many years ago, when photographers needed to adjust a photo, they’d have to spend time in the darkroom tinkering with equipment and chemicals. Then along came digital imagery and the magic of Photoshop, which brought photo manipulation out of the dark. Nowadays, smartphone apps can manipulate photos in myriad ways.

Photoshop is the granddaddy of photo manipulation on a smartphone. The Photoshop Touch app is a touch-screen version of the original. It’s powerful and can apply many of the effects you have seen in magazines and on the Web. That includes adjustments like the tilt-shift effect, which can make a townscape look like a miniature model, as well as better-known
Not so many years ago, when photographers needed to adjust a photo, they'd have to spend time in the darkroom tinkering with equipment and chemicals. Then along came digital imagery and the magic of Photoshop which brought photo manipulation out of the dark. Nowadays, smartphone apps can manipulate photos in myriad ways. Photoshop is the granddaddy of photo manipulation on a smartphone. The Photoshop Touch app is a touch-screen version of the original. It's powerful and can apply many of the effects you have seen in magazines and on the Web. That includes adjustments like the tilt-shift effect which can make a townscape look like a miniature model as well as better-known.
Singleton vs. Coreferent (entities)

on OntoNotes CoNLL-2012 dev
Singleton vs. Coreferent (mentions) on OntoNotes CoNLL-2012 dev
Can we tease singletons apart?

Is a mention likely to be part of coreference chain or not?

Useful insights from the dynamic semantics literature

Will I live much longer?
Outline

1. Linguistic insights: Lifespan features
2. Data
3. Lifespan model results
4. Coreference resolution results
Lifespan model: logistic regression

What is the probability that a mention is coreferent?
How is the lifespan of discourse entities shaped by its mention features?

Local morphosyntatic features

But also features of the syntactic and semantic environment in which mentions occur
Features of the lifespan model

Morphologic, syntactic and semantic cues to predict the lifespan of entities

Explicitly use linguistic insights about scope of modals, negation and attitude predicates

Interactions between scope and mention type
Interactions with semantic operators

[i.a., Asher, Bittner, Harris, Heim, Israel, Kamp, Karttunen, Ladusaw, Lascarides, Roberts, Stone]

Indefinites excel at introducing long-lived discourse entities:

\[ S \]

\[ \text{Kim} \quad \text{VP} \]

\[ \text{skipped} \quad \text{NP} \]

\[ \text{an exam question}. \]

\[ \checkmark \text{It was too hard}. \]
Negation often blocks anaphora for indefinites

[Karttunen 1976]

#It has a long string.
Modality often blocks anaphora for indefinites

[Karttunen 1976]

#It has a long string.
Quantificational operators disfavor anaphora

\[
S \\
NP  \\
\text{Everyone}
\]

\[
VP \\
\text{took} \\
NP \\
a \text{different exam.}
\]

\[\text{It was too hard.}\]
Non-factive attitude predicates disfavor anaphora

[SandyVPclaims that Jesse bought a bike.]

[John wants S to catch NP a fish.]

#Do you see it from here?

[Karttunen 1976]
Non-factive attitude predicates disfavor anaphora

[Sandy]

S

VP

claims

SBAR

that Jesse bought

NP

a bike.

#It has a green frame.

[Karttunen 1976]
Pragmatic veridicality facilitates anaphora

[S] [NP] [VP] [S] [that] [NP] [passed.]

reported

[It was of course controversial.]

Internal morphosyntactic features

[i.a., Aissen, Asher, Hovy, Marcus, McCready, Palmer, Prince, Partee, Schwarzschild, Wang]

Features standardly used in coreference systems:

- Type
- Animacy
- Person
- Number
- Quantification/Definiteness
- Number of modifiers
- Named entity
Syntactic features

[i.a., Beaver, Birner, Chambers, Chang, Chafe, Grosz, Hobbs, Joshi, Jurafsky, Lee, Peirsman, Prince, Surdeanu, Walker, Ward, Weinstein]

From work in Centering Theory and Information Structuring, coreferent mentions:
- will favor sentence-initial (topic-tracking) positions
- are likely to appear as core verbal arguments

Features:
- grammatical relation (subject, object, noun modifier, …)
- sentence position
- in coordination
Bills on ports and immigration were submitted by Senator Brownback
Semantic environments

Under the scope of

- negation
- modal
- attitude predicate

+ interactions with mention type
definiteness
## Data: OntoNotes CoNLL-2012 (English)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Docs</th>
<th>Tokens</th>
<th>Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coreferent</td>
</tr>
<tr>
<td>Training</td>
<td>2,802</td>
<td>1.3M</td>
<td>152,828</td>
</tr>
<tr>
<td>Dev</td>
<td>343</td>
<td>160K</td>
<td>18,815</td>
</tr>
<tr>
<td>Test</td>
<td>348</td>
<td>170K</td>
<td>19,392</td>
</tr>
</tbody>
</table>
## Results corroborate linguistic insights

<table>
<thead>
<tr>
<th>Favor coreference</th>
<th>Disfavor coreference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>Inanimate</td>
</tr>
<tr>
<td>Singular</td>
<td>Plural/Mass</td>
</tr>
<tr>
<td>Proper name/NE</td>
<td>Quantified</td>
</tr>
<tr>
<td>Pronominal</td>
<td><strong>Indefinite</strong></td>
</tr>
</tbody>
</table>

Intuitive except for indefinites. They are supposed to introduce discourse referents...
Lifespan model: morphosyntax

Named entity:

We used the 18 OntoNotes types

All favor coreference except

- ORDINAL
- PERCENT
- MONEY
- QUANTITY
Results confirm findings of Centering Theory and Information Structuring

<table>
<thead>
<tr>
<th>Favor coreference</th>
<th>Disfavor coreference</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentence-initial</td>
<td>sentence medial or final</td>
</tr>
<tr>
<td>subject or object</td>
<td>non-core argument</td>
</tr>
<tr>
<td></td>
<td>in coordination</td>
</tr>
</tbody>
</table>
## Semantic environments

<table>
<thead>
<tr>
<th>Favor coreference</th>
<th>Disfavor coreference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under attitude verb</td>
<td>Under negation</td>
</tr>
<tr>
<td></td>
<td>Under modality</td>
</tr>
</tbody>
</table>
## Interactions: scope and mention type

<table>
<thead>
<tr>
<th>Favor coreference</th>
<th>Disfavor coreference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negation * pronoun/name</td>
<td>Negation * indefinite</td>
</tr>
<tr>
<td>Modal * pronoun/name</td>
<td>Modal * indefinite</td>
</tr>
<tr>
<td>Attitude V * pronoun/name</td>
<td></td>
</tr>
<tr>
<td>Attitude V * indefinite</td>
<td></td>
</tr>
</tbody>
</table>
Lifespan model is effective by itself on OntoNotes CoNLL-2012 dev

<table>
<thead>
<tr>
<th></th>
<th>Singleton</th>
<th>Coreferent</th>
<th>Accuracy</th>
<th># Feat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>LINGUISTIC</td>
<td>80.2</td>
<td>77.5</td>
<td>71.4</td>
<td>74.6</td>
</tr>
</tbody>
</table>
Lifespan model is effective by itself on OntoNotes CoNLL-2012 dev

<table>
<thead>
<tr>
<th></th>
<th>Singleton</th>
<th></th>
<th>Coreferent</th>
<th></th>
<th>Accuracy</th>
<th># Feat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINGUISTIC</td>
<td>80.2</td>
<td>77.5</td>
<td>71.4</td>
<td>74.6</td>
<td>76.3</td>
<td>123</td>
</tr>
<tr>
<td>SURFACE</td>
<td>80.2</td>
<td>79.9</td>
<td>75.3</td>
<td>75.6</td>
<td>78.0</td>
<td>73,393</td>
</tr>
</tbody>
</table>
Lifespan model is effective by itself on OntoNotes CoNLL-2012 dev

<table>
<thead>
<tr>
<th></th>
<th>Singleton</th>
<th>Coreferent</th>
<th>Accuracy</th>
<th># Feat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>LINGUISTIC</td>
<td>80.2</td>
<td>77.5</td>
<td>71.4</td>
<td>74.6</td>
</tr>
<tr>
<td>SURFACE</td>
<td>80.2</td>
<td>79.9</td>
<td>75.3</td>
<td>75.6</td>
</tr>
<tr>
<td>COMBINED</td>
<td>81.1</td>
<td>80.9</td>
<td>76.4</td>
<td>76.6</td>
</tr>
</tbody>
</table>
**Lifespan model is effective by itself**

on OntoNotes CoNLL-2012 dev

<table>
<thead>
<tr>
<th></th>
<th>Singleton</th>
<th>Coreferent</th>
<th>Accuracy</th>
<th># Feat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>LINGUISTIC</td>
<td>80.2</td>
<td>77.5</td>
<td>71.4</td>
<td>74.6</td>
</tr>
<tr>
<td>SURFACE</td>
<td>80.2</td>
<td>79.9</td>
<td>75.3</td>
<td>75.6</td>
</tr>
<tr>
<td>COMBINED</td>
<td>81.1</td>
<td>80.9</td>
<td>76.4</td>
<td>76.6</td>
</tr>
<tr>
<td>CONFIDENT</td>
<td>56.0</td>
<td>89.8</td>
<td>48.2</td>
<td>90.7</td>
</tr>
<tr>
<td></td>
<td><strong>Pr &lt; 0.2</strong></td>
<td><strong>Pr &gt; 0.8</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stanford coreference system

[Lee et al. 2011]
Stanford coreference system

Remove 33% of the mentions

[Lee et al. 2011]
Recall drops, precision increases

Stanford w/ Lifespan model

MUC

B3

CEAF-m

CEAF-e

BLANC
<table>
<thead>
<tr>
<th><strong>President Clinton</strong> is questioning the legitimacy of G. W. Bush’s election victory. Speaking last night to Democratic supporters in Chicago, <strong>he</strong> said Bush won the election only because Republicans stopped the vote-counting in Florida, and <strong>Mr. Clinton</strong> praised Al Gore’s campaign manager, Bill Daley, for the way he handled the election. “<strong>I</strong> want to thank Bill Daley for his exemplary service as Secretary of Commerce. He was brilliant. <strong>I</strong> think he did a brilliant job in leading Vice President Gore to victory <strong>myself.”</strong></th>
<th><strong>7 error types:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Missing entities</td>
<td>- Missing mentions</td>
</tr>
<tr>
<td>- Extra entities</td>
<td>- Extra mentions</td>
</tr>
<tr>
<td>- Conflated entities</td>
<td>- Span errors</td>
</tr>
<tr>
<td>- Divided entities</td>
<td></td>
</tr>
</tbody>
</table>
Stanford coref system’s errors

Conflated entities
- Stanford alone: 1635
- with lifespan: 1607

Extra entity
- Stanford alone: 897
- with lifespan: 728

Extra mention
- Stanford alone: 535
- with lifespan: 523

Divided entities
- Stanford alone: 2038
- with lifespan: 2021

Missing entity
- Stanford alone: 830
- with lifespan: 877

Missing mention
- Stanford alone: 1154
- with lifespan: 1158
Berkeley coreference system

\[ P(a|x) \propto \exp \left( \sum_{i=1}^{n} w^\top f(i, a_i, x) \right) \]

[Voters]_1 agree when [they]_1 are given [a chance]_2 to decide if [they]_1 ...
Berkeley coreference system

Learning system  

We add to mentions their singleton/coreferent probability as given by the “combined” lifespan model  

[Durrett & Klein 2013]
Recall (often) increases, precision increases

Berkeley w/ Lifespan model

MUC

B3

CEAF-m

CEAF-e

BLANC
Berkeley coref system's errors

Conflated entities

Extra entity

Extra mention

Divided entities

Missing entity

Missing mention
Lifespan improves coreference resolution

The differences are statistically significant.
New perspective on bridging

I looked into the room. *The ceiling* was very high.

singleton anchor

I looked into the room. *It* was large. *The ceiling* was high.

50 documents of OntoNotes annotated for bridging

[145 are singleton anchors]

[Hou et al. 2013]
Lifespan can model bridging

Dataset of 145 singleton anchors and 145 randomly sampled true singletons

Linguistic model obtains a mean F1 of 65%

Singleton anchors are different from true singletons in a way that our features are able to capture

Bridging is governed by its own mix of linguistic and contextual factors [Hawkins 1978, Prince 1981, Schwarz 2009]
Wrapping up

Use NLP data to evaluate linguistic theories

Fact: Large proportion of singleton entities

Lifespan model
- tells apart singletons and coreferent mentions
- corroborates linguistic insights
- improves coreference resolution