Scratching the Surface of Possible Translations

Ondřej Bojar Matouš Macháček Aleš Tamchyna Daniel Zeman

Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics (ÚFAL)

TSD 2013

Outline

1. Introduction
2. Compact Representations for Many References
3. Collected Data
4. Analysis
5. Conclusion

Introduction

Introduction

Broad perspective:

- Relation between sentence meaning and the string is complicated.
  - Many variations of a sentence preserve the meaning.
  - A little change in a word can drastically reverse the meaning.
- We want a tangible dataset that represents this relation.

Introduction

Broad perspective:

- Relation between sentence meaning and the string is complicated.
  - Many variations of a sentence preserve the meaning.
  - A little change in a word can drastically reverse the meaning.
- We want a tangible dataset that represents this relation.

Narrow use case:

- Machine translation is evaluated by comparing output to a reference.
  - There are many good translations of a sentence
  - Only very few of them are available as a reference translation.
- We want to improve MT evaluation using more/all references.

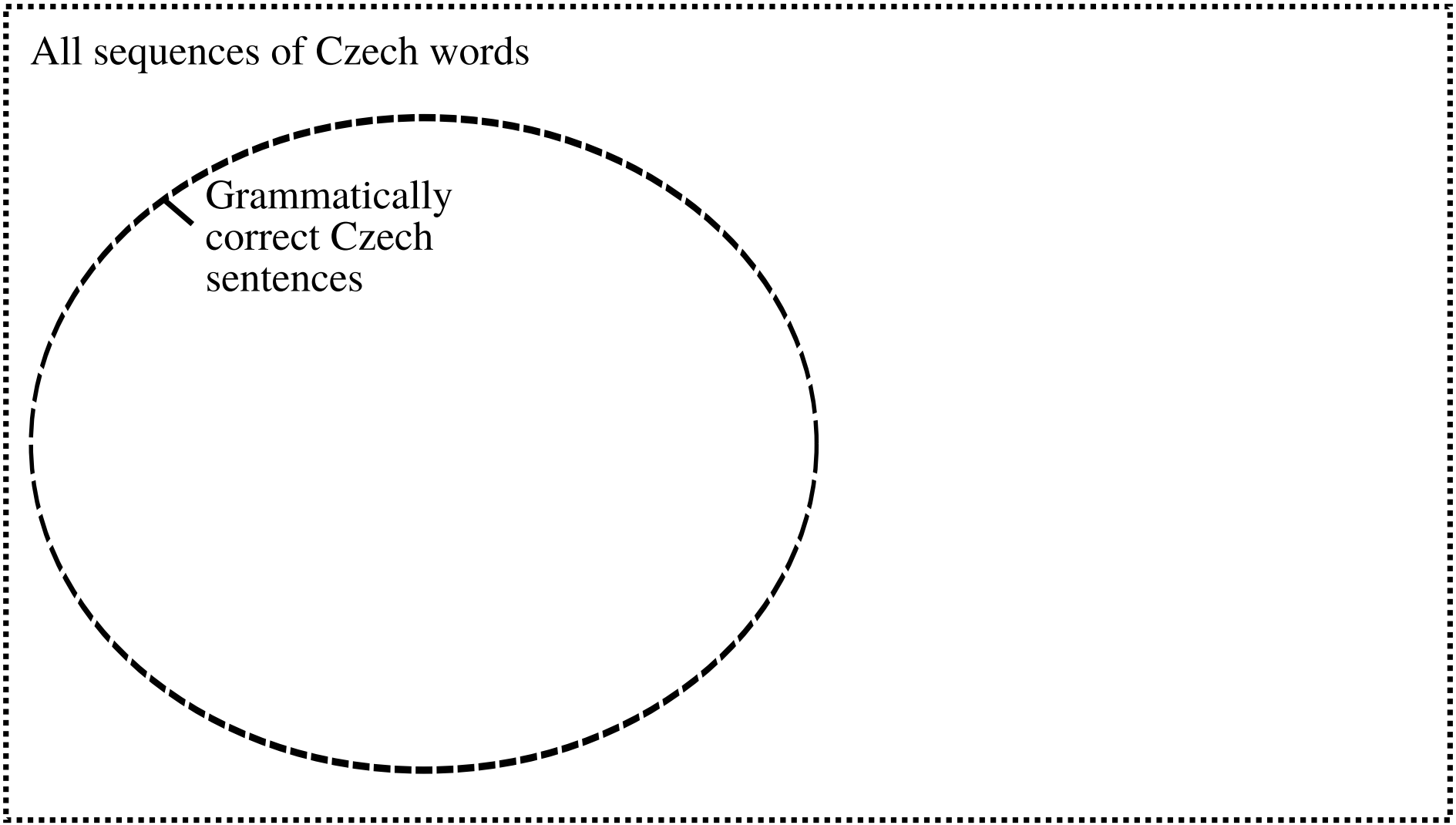
Space of Possible Translations

For a single English source sentence, we can consider:

All sequences of Czech words

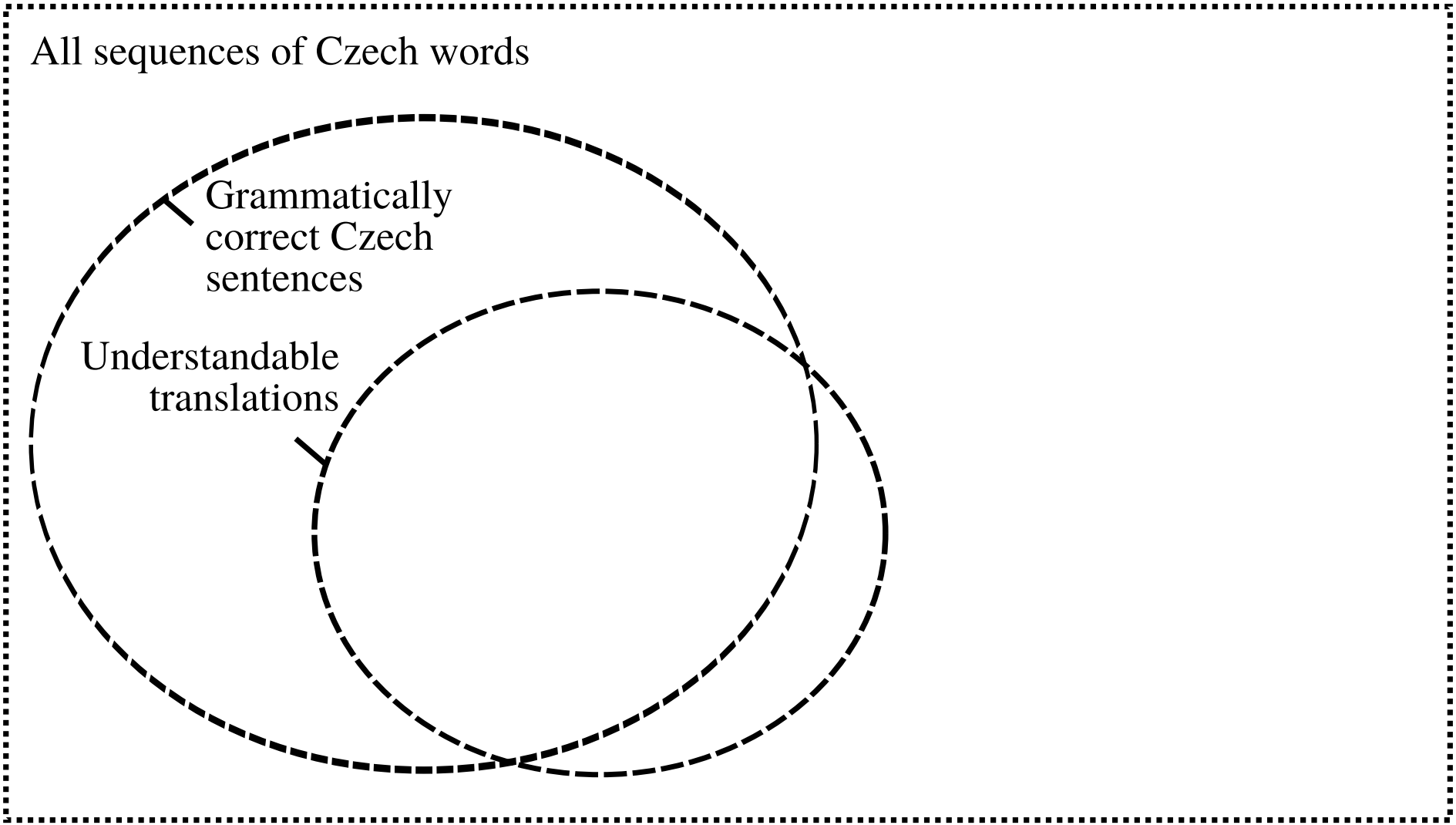
Space of Possible Translations

For a single English source sentence, we can consider:



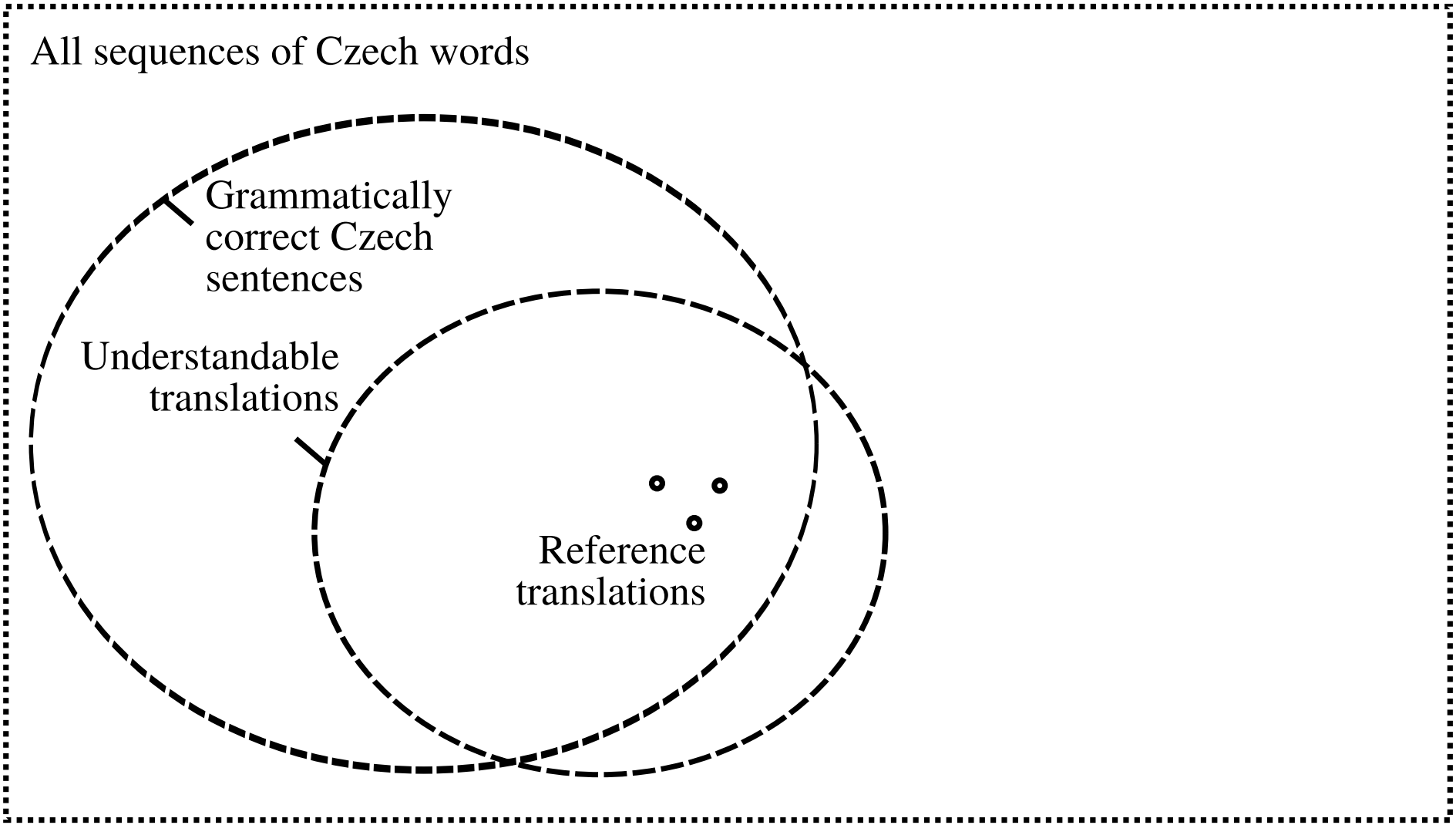
Space of Possible Translations

For a single English source sentence, we can consider:



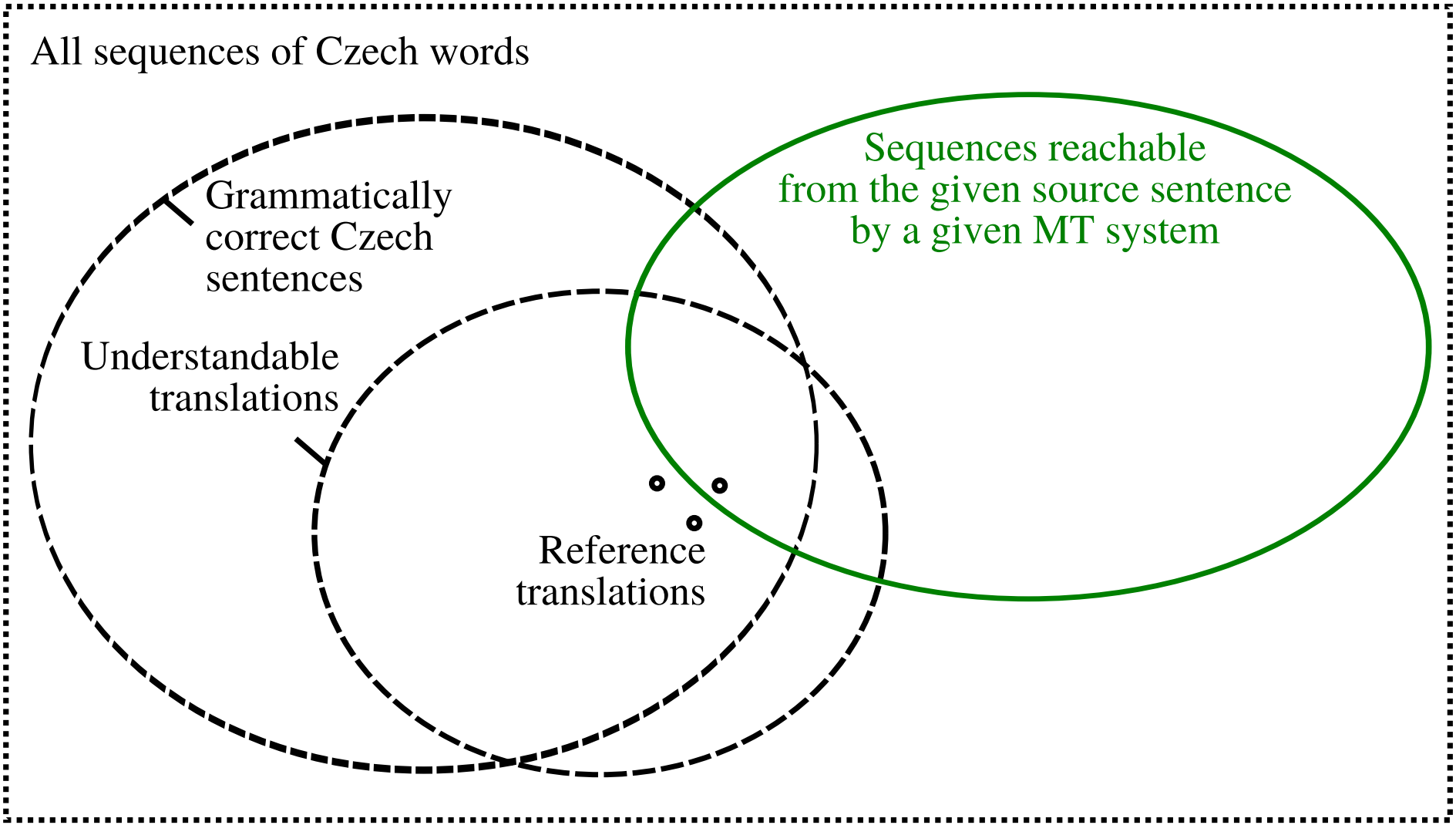
Space of Possible Translations

For a single English source sentence, we can consider:



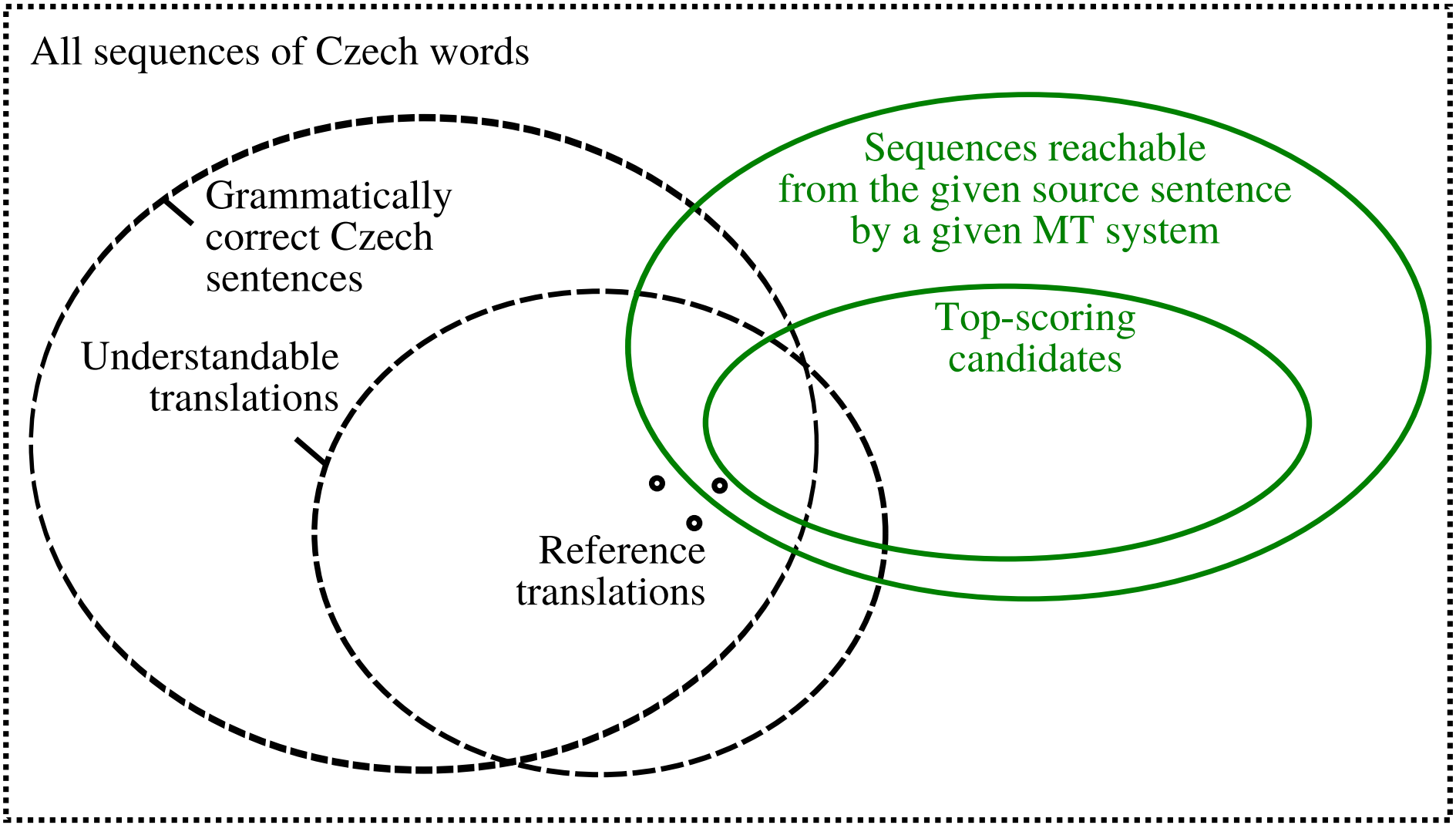
Space of Possible Translations

For a single English source sentence, we can consider:



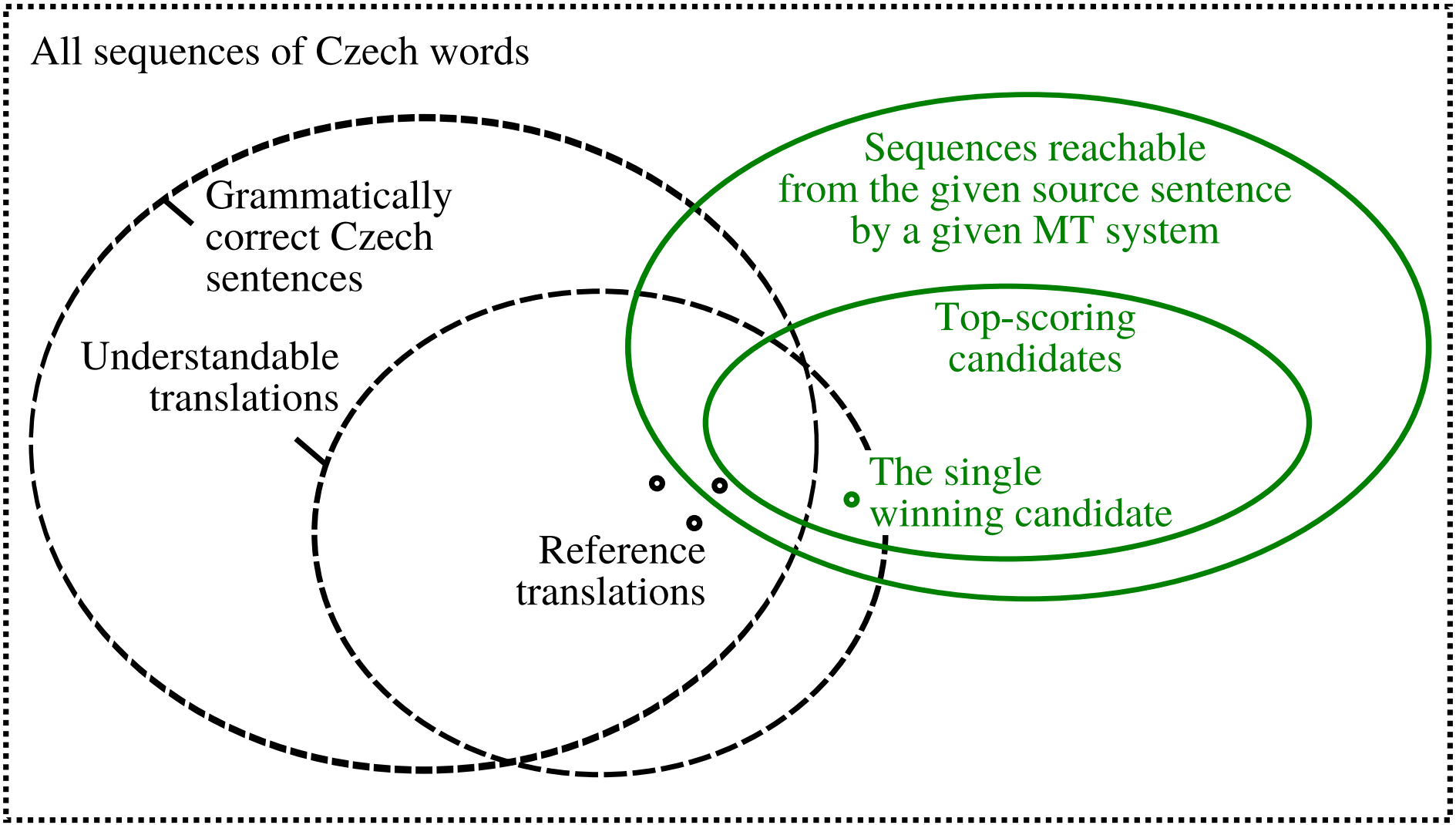
Space of Possible Translations

For a single English source sentence, we can consider:



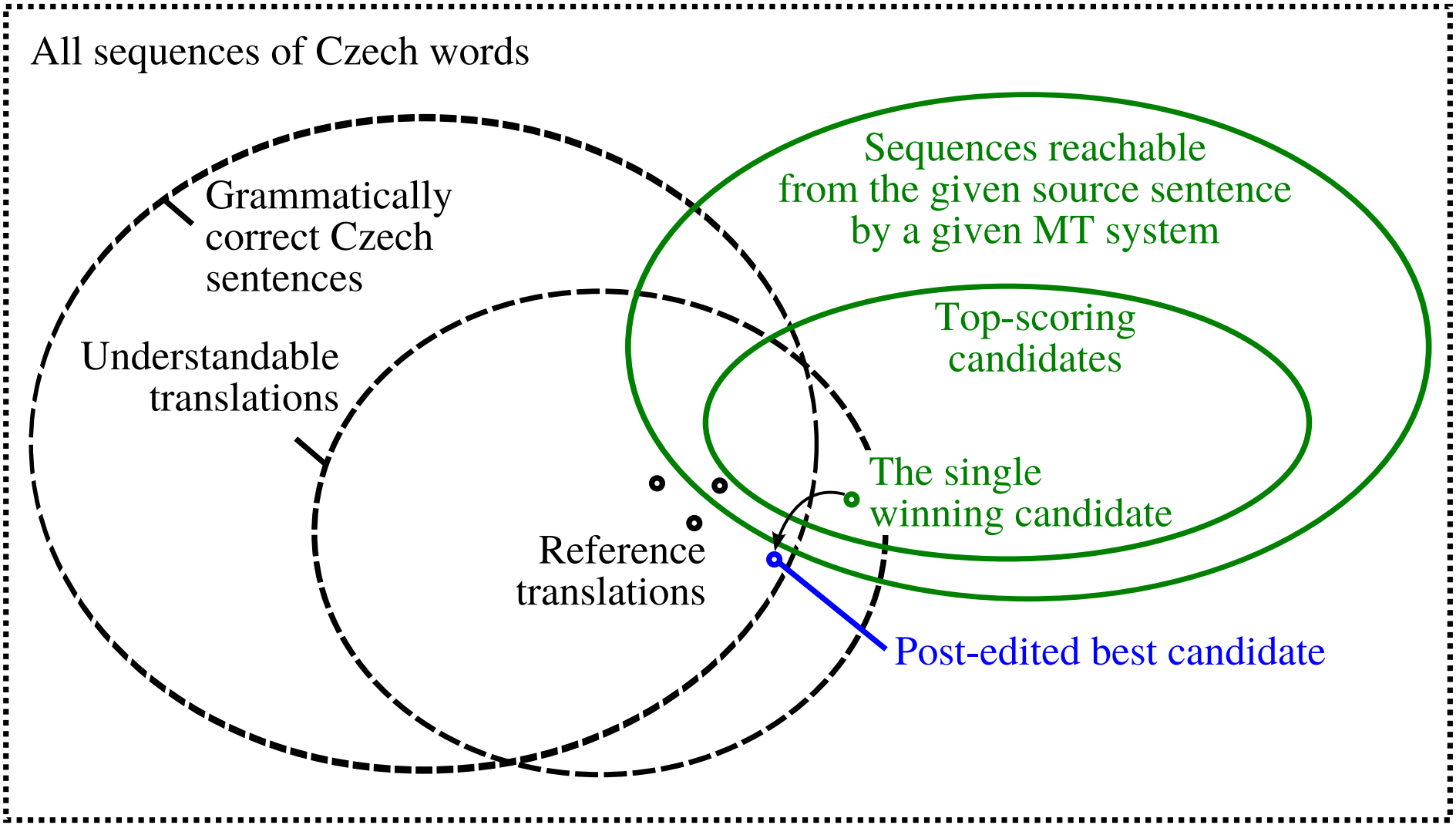
Space of Possible Translations

For a single English source sentence, we can consider:



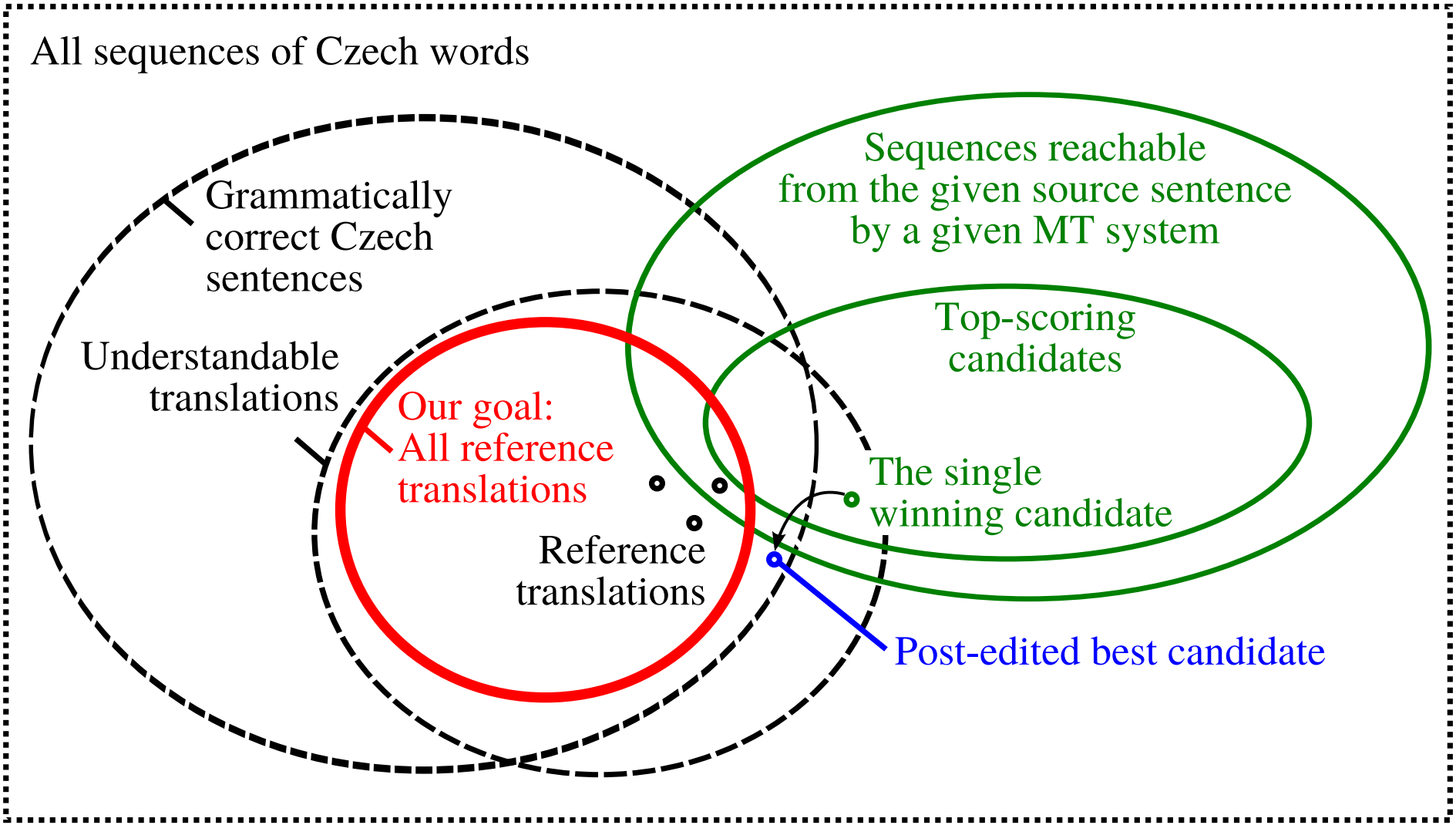
Space of Possible Translations

For a single English source sentence, we can consider:



Space of Possible Translations

For a single English source sentence, we can consider:



Compact Representation of Many References

- Typing out all references is not possible.

Compact Representation of Many References

- Typing out all references is not possible.
- Dreyer and Marcu (2012) use Recursive Transition Networks (RTN).

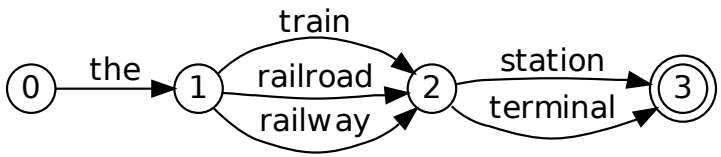
Compact Representation of Many References

- Typing out all references is not possible.
- Dreyer and Marcu (2012) use Recursive Transition Networks (RTN).
- An RTN consists of **cards**:

Compact Representation of Many References

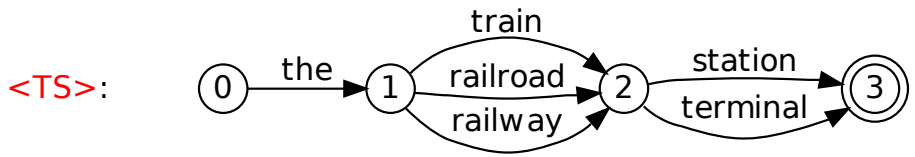
- Typing out all references is not possible.
- Dreyer and Marcu (2012) use Recursive Transition Networks (RTN).
- An RTN consists of **cards**:
  - A card covers multiple translations of a short phrase:

<TS>:

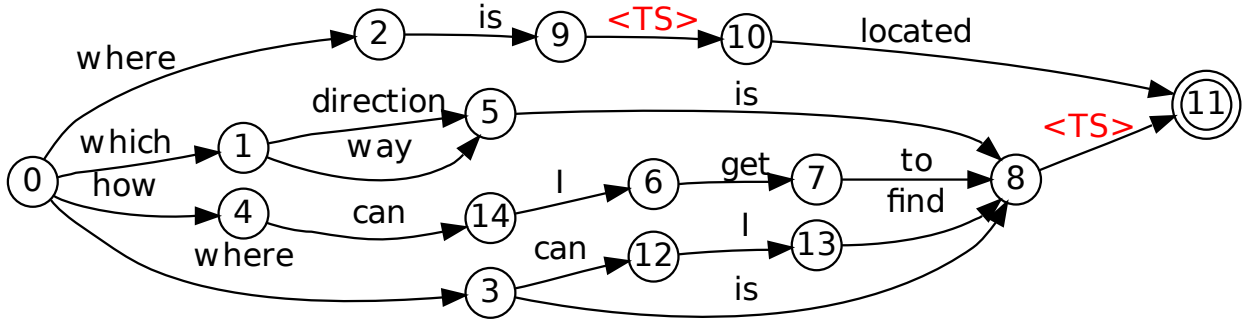


Compact Representation of Many References

- Typing out all references is not possible.
- Dreyer and Marcu (2012) use Recursive Transition Networks (RTN).
- An RTN consists of **cards**:
  - A card covers multiple translations of a short phrase:



  - Cards are reused within other cards to built the sentence:



Recursive Transition Networks for Czech

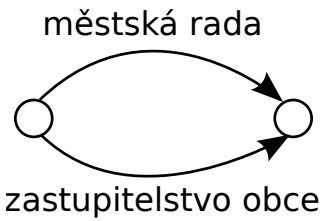
The city council approved a new regulation.

Recursive Transition Networks for Czech

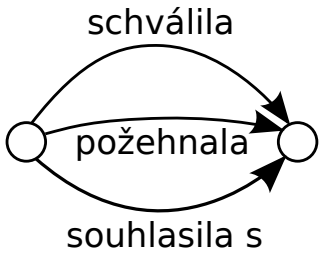
The city council approved a new regulation

Recursive Transition Networks for Czech

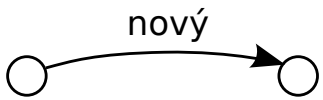
The city council



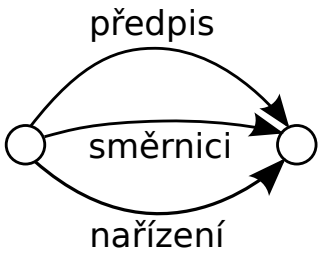
approved



a new

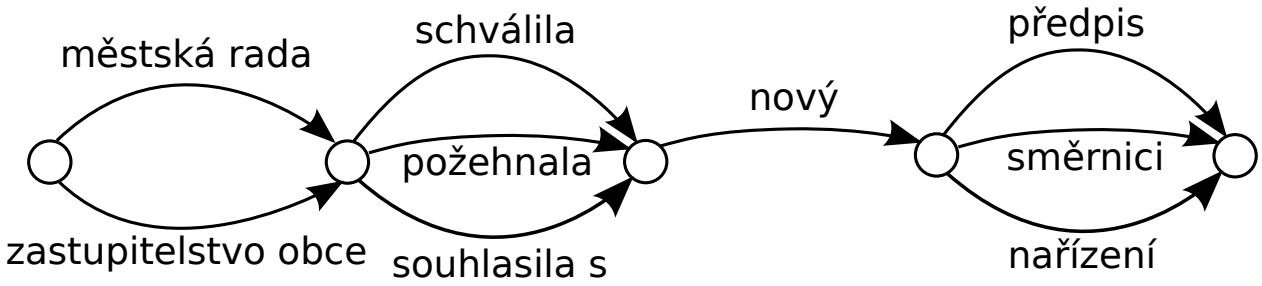


regulation



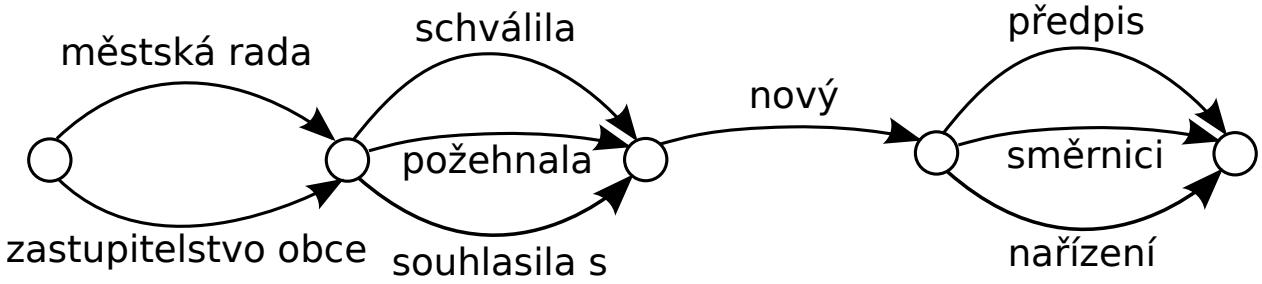
Recursive Transition Networks for Czech

The city council approved a new regulation.



Recursive Transition Networks for Czech

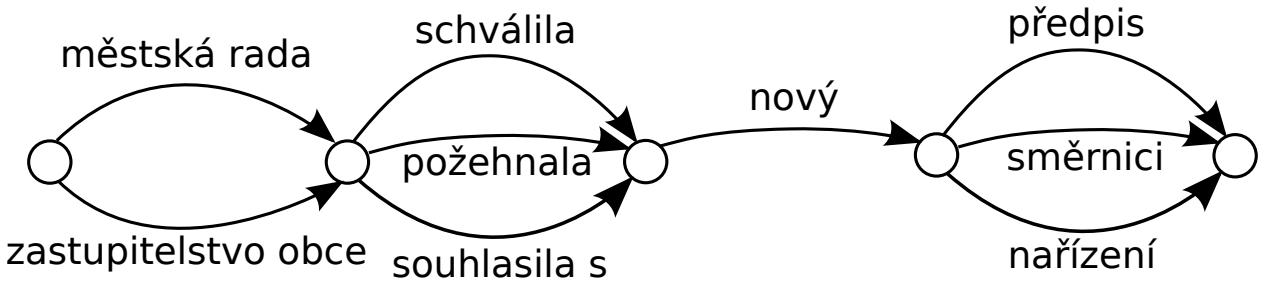
The city council approved a new regulation.



Městská rada schválila nový předpis

Recursive Transition Networks for Czech

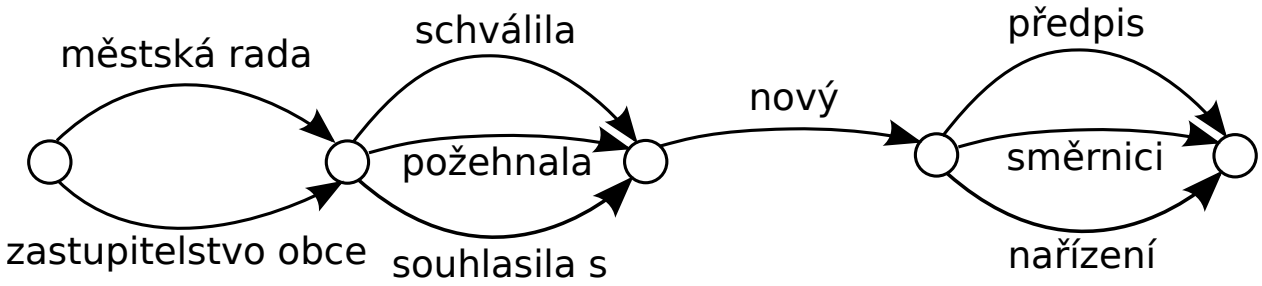
The city council approved a new regulation.



Městská rada schválila nový předpis ✔

Recursive Transition Networks for Czech

The city council approved a new regulation.

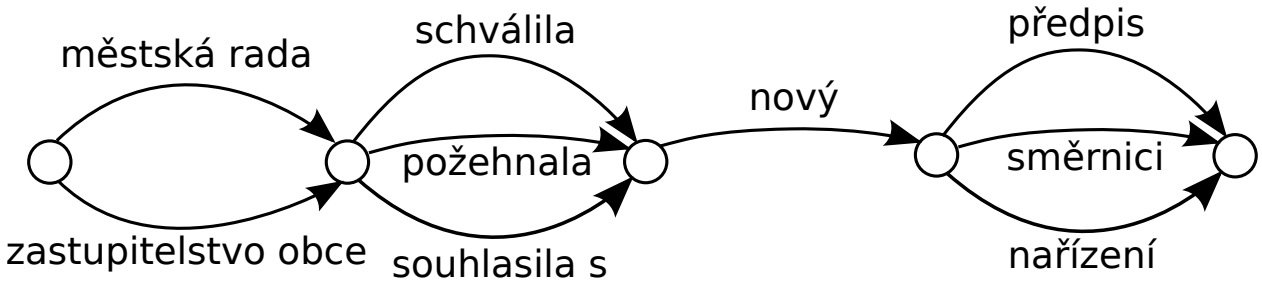


Městská rada schválila nový předpis ✔

Zastupitelstvo$_{\text{neut}}$ obce schválila$_{\text{fem}}$ nový předpis

Recursive Transition Networks for Czech

The city council approved a new regulation.

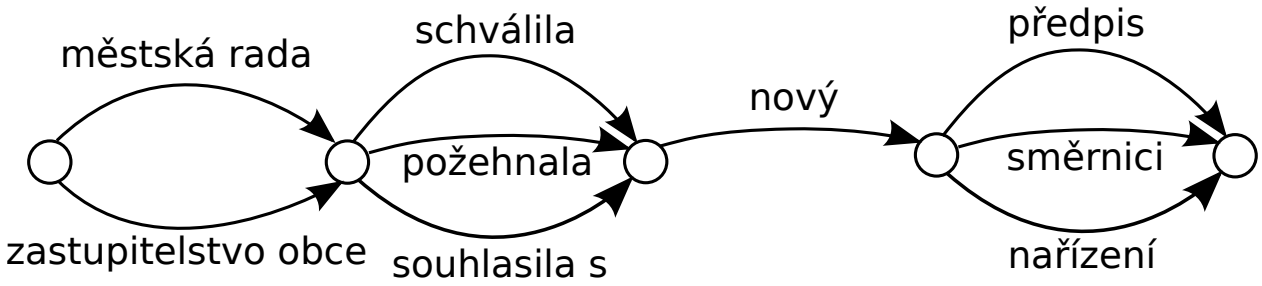


Městská rada schválila nový předpis ✔

Zastupitelstvo$_{neut}$ obce schválila$_{fem}$ nový předpis ✖

Recursive Transition Networks for Czech

The city council approved a new regulation.



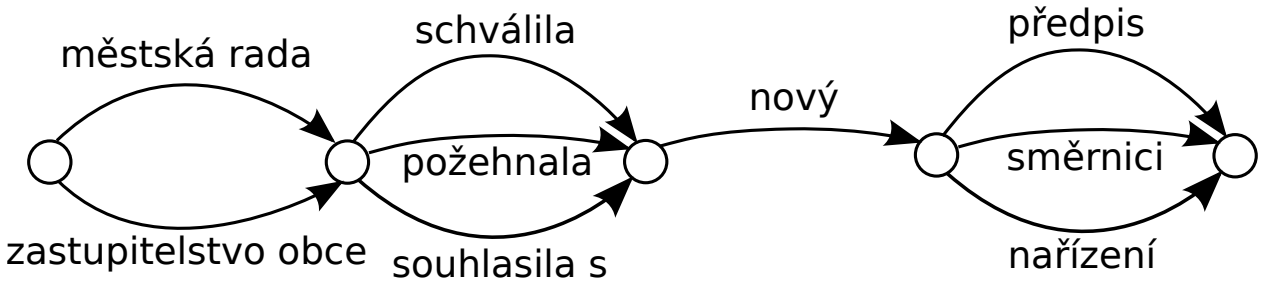
Městská rada schválila nový předpis ✔

Zastupitelstvo$_{neut}$ obce schválila$_{fem}$ nový předpis ✖

Městská rada schválila nový$_{masc}$ směrnici$_{fem}$

Recursive Transition Networks for Czech

The city council approved a new regulation.



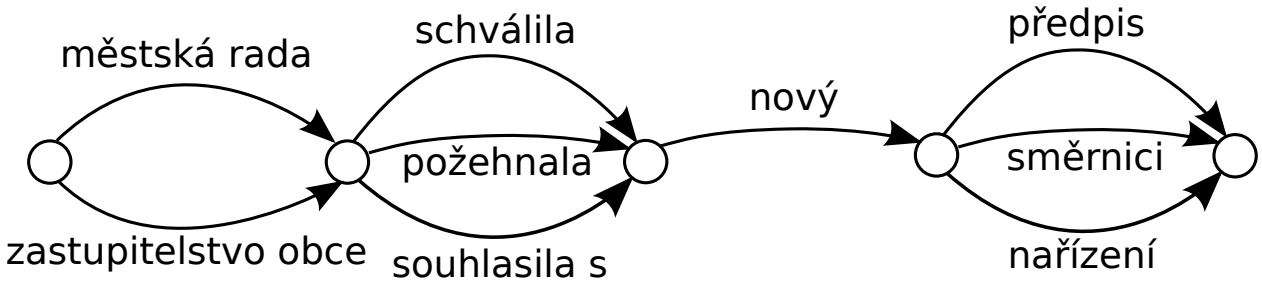
Městská rada schválila nový předpis ✔

Zastupitelstvo$_{neut}$ obce schválila$_{fem}$ nový předpis ✖

Městská rada schválila nový$_{masc}$ směrnici$_{fem}$ ✖

Recursive Transition Networks for Czech

The city council approved a new regulation.



Městská rada schválila nový předpis ✔

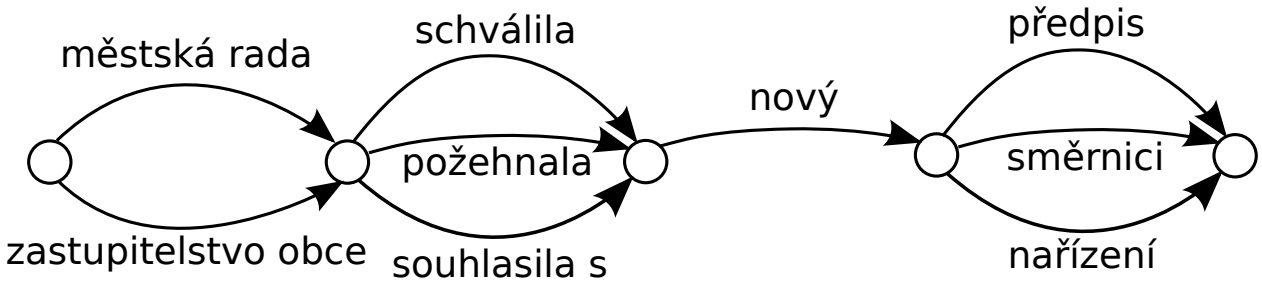
Zastupitelstvo$_{\text{neut}}$ obce schválila$_{\text{fem}}$ nový předpis ✖

Městská rada schválila nový$_{\text{masc}}$ směrnici$_{\text{fem}}$ ✖

Městská rada souhlasila$_{\text{requires ins}}$ nový předpis$_{\text{nom}}$

Recursive Transition Networks for Czech

The city council approved a new regulation.



Městská rada schválila nový předpis ✔

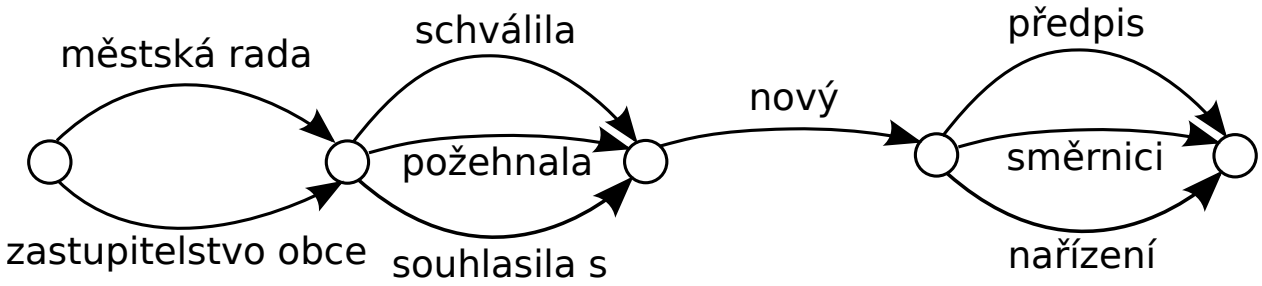
Zastupitelstvo$_{\text{neut}}$ obce schválila$_{\text{fem}}$ nový předpis ✖

Městská rada schválila nový$_{\text{masc}}$ směrnici$_{\text{fem}}$ ✖

Městská rada souhlasila$_{\text{requires ins}}$ nový předpis$_{\text{nom}}$ ✖

Recursive Transition Networks for Czech

The city council approved a new regulation.



Městská rada schválila nový předpis ✔

Zastupitelstvo$_{\text{neut}}$ obce schválila$_{\text{fem}}$ nový předpis ✖

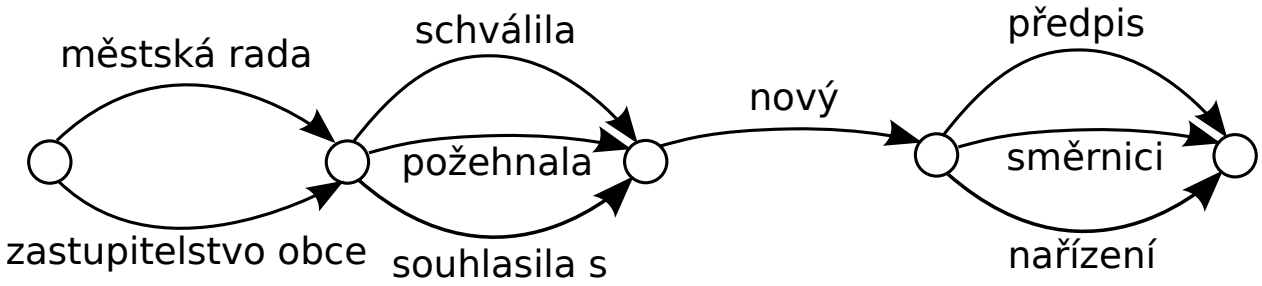
Městská rada schválila nový$_{\text{masc}}$ směrnici$_{\text{fem}}$ ✖

Městská rada souhlasila$_{\text{requires ins}}$ nový předpis$_{\text{nom}}$ ✖

- Czech needs to specify constraints under which cards can combine.

Recursive Transition Networks for Czech

The city council approved a new regulation.



Městská rada schválila nový předpis ✔

Zastupitelstvo$_{\text{neut}}$ obce schválila$_{\text{fem}}$ nový předpis ✖

Městská rada schválila nový$_{\text{masc}}$ směrnici$_{\text{fem}}$ ✖

Městská rada souhlasila$_{\text{requires ins}}$ nový předpis$_{\text{nom}}$ ✖

- Czech needs to specify constraints under which cards can combine.
- The workaround of using many card types would be cumbersome.

Unification-Based Annotation

- We proposed a unification-based representation suitable for Czech.

Unification-Based Annotation

- We proposed a unification-based representation suitable for Czech.
- Main building block called bubble, defined by:

Unification-Based Annotation

- We proposed a unification-based representation suitable for Czech.
- Main building block called bubble, defined by:
  - The set of source language tokens it covers

Unification-Based Annotation

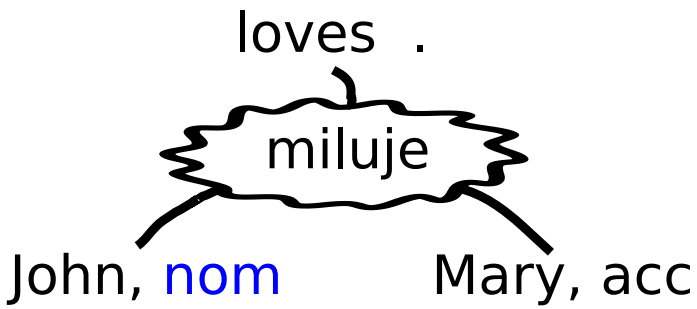
- We proposed a unification-based representation suitable for Czech.
- Main building block called bubble, defined by:
  - The set of source language tokens it covers
  - The set of conditions it meets

Unification-Based Annotation

- We proposed a unification-based representation suitable for Czech.
- Main building block called bubble, defined by:
  - The set of source language tokens it covers
  - The set of conditions it meets
  - A translation alternative, composed of:
    - atoms – tokens of the target language
    - slots – positions with given conditions which bubbles must fulfil to fill the slot

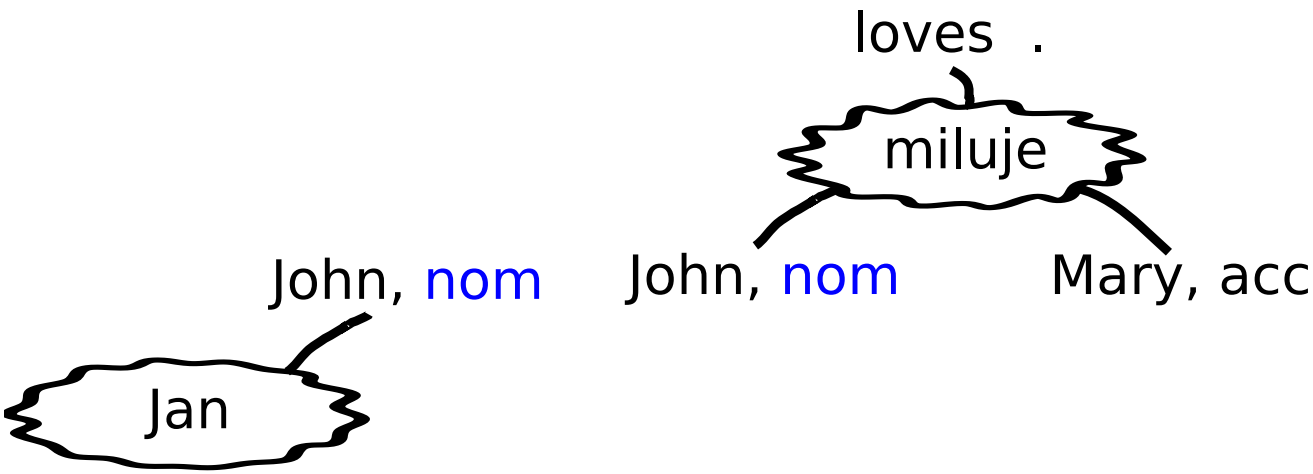
Bubble Annotation

- Input: John loves Mary.
- Annotation = bubbles covering parts of input.



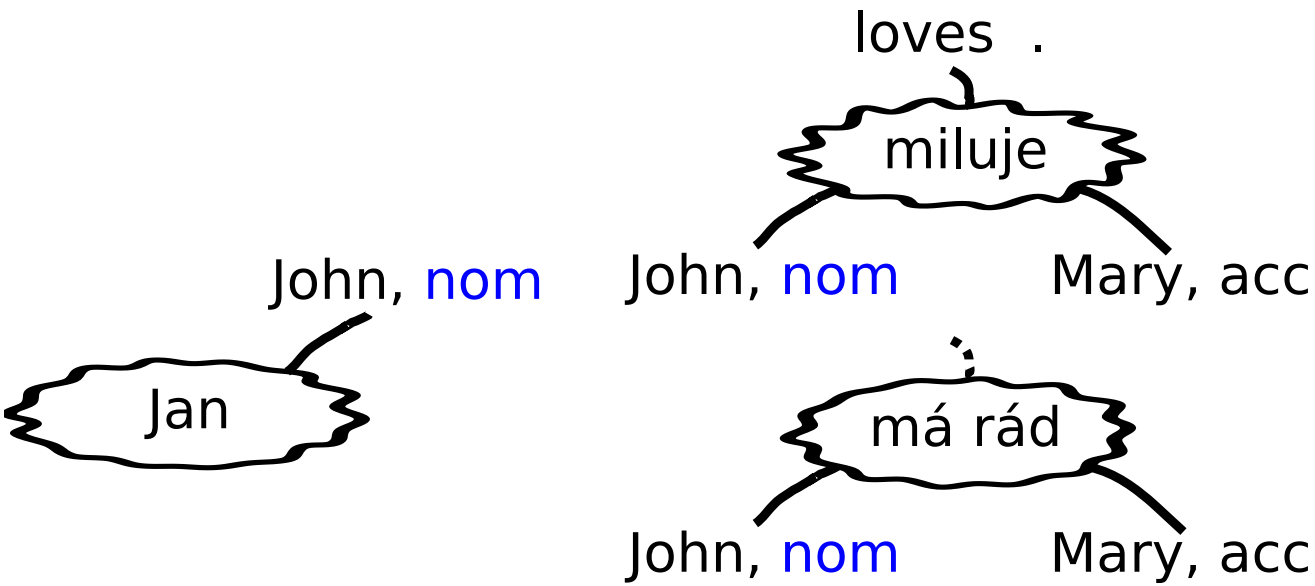
Bubble Annotation

- Input: John loves Mary.
- Annotation = bubbles covering parts of input.



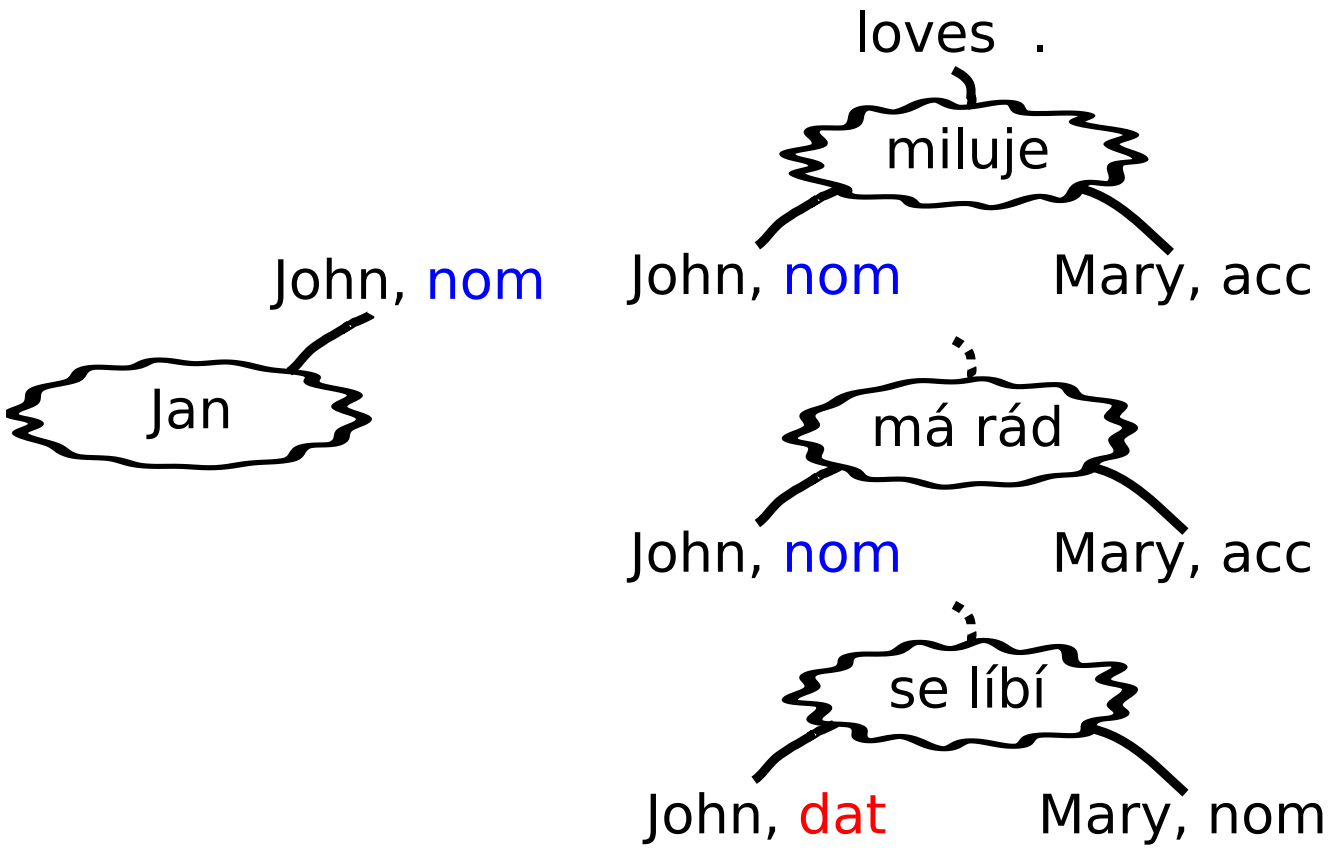
Bubble Annotation

- Input: John loves Mary.
- Annotation = bubbles covering parts of input.



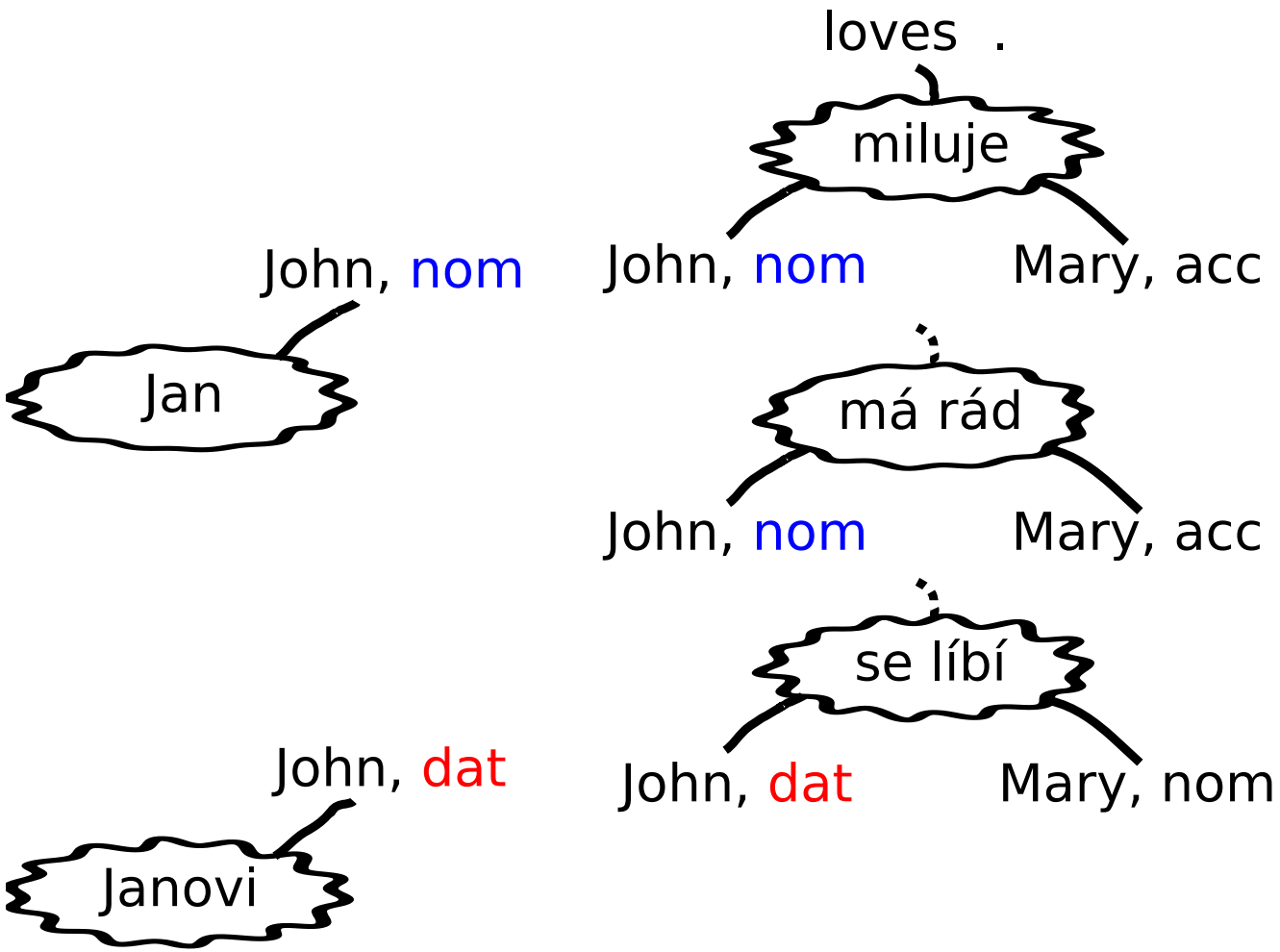
Bubble Annotation

- Input: John loves Mary.
- Annotation = bubbles covering parts of input.



Bubble Annotation

- Input: John loves Mary.
- Annotation = bubbles covering parts of input.



Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.

Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

Annotating Translations: Prolog Interface

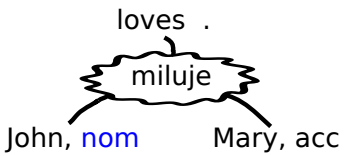
- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

```
option( this bubble covers , this bubble safisfies , this bubble provides and requires ).
```

Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

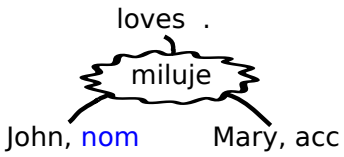
`option(` this bubble covers `,` this bubble safisfies `,` this bubble provides and requires `).`



Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

`option(` this bubble covers `,` this bubble safisfies `,` this bubble provides and requires `).`

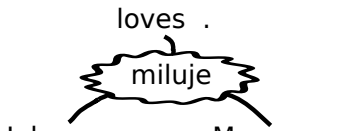



```
option( [loves, .], [],
        [ [John, nom],
          miluje,
          [Mary, acc] ] ).
```

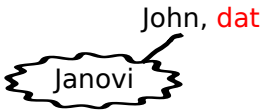
Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

`option(` this bubble covers `,` this bubble safisfies `,` this bubble provides and requires `).`



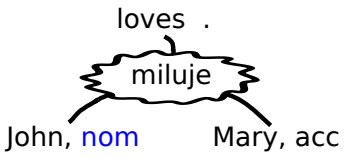

```
option( [loves, .], [],
        [ [John, nom],
          miluje,
          [Mary, acc] ] ).
```



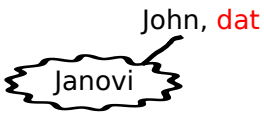
Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

`option(` this bubble covers `,` this bubble safisfies `,` this bubble provides and requires `).`




```
option( [loves, .], [],
        [ [John, nom],
          miluje,
          [Mary, acc] ] ).
```

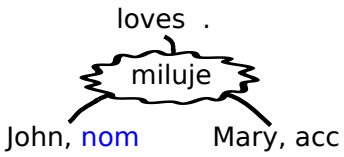



```
option( [John], [dat], [Janovi] ).
```

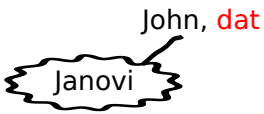
Annotating Translations: Prolog Interface

- Prolog is ideal for constraints evaluation and bubble expansion.
- The annotator specifies a set of bubbles as clauses for `option/2`.

```
option( this bubble covers , this bubble safisfies , this bubble provides and requires ).
```




```
option( [loves, .], [], [ [John, nom], miluje, [Mary, acc] ] ).
```




```
option( [John], [dat], [Janovi] ).
```

- About 300 lines of a common code generates all possible sentences.

Web Interface

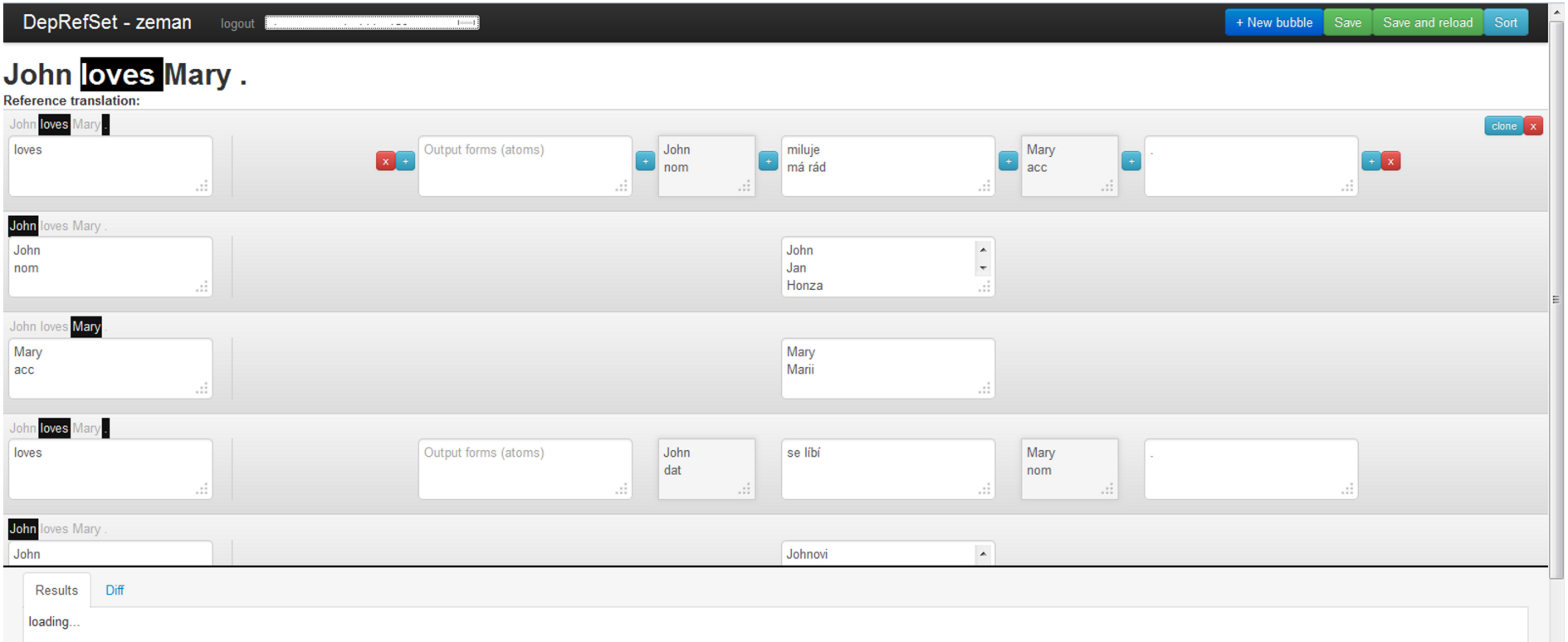
- Editing prolog source code is not for everyone.

Web Interface

- Editing prolog source code is not for everyone.
- Our web-based graphical interface hides away the Prolog syntax:

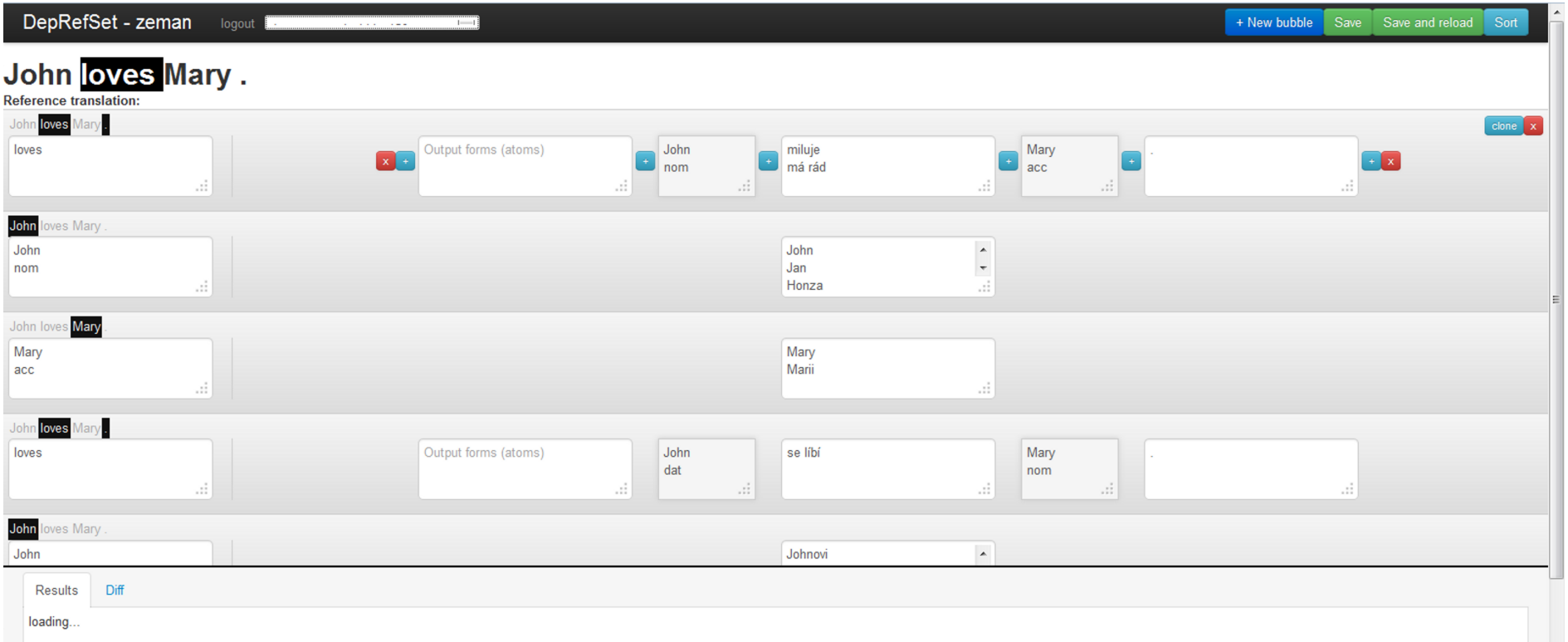
Web Interface

- Editing prolog source code is not for everyone.
- Our web-based graphical interface hides away the Prolog syntax:

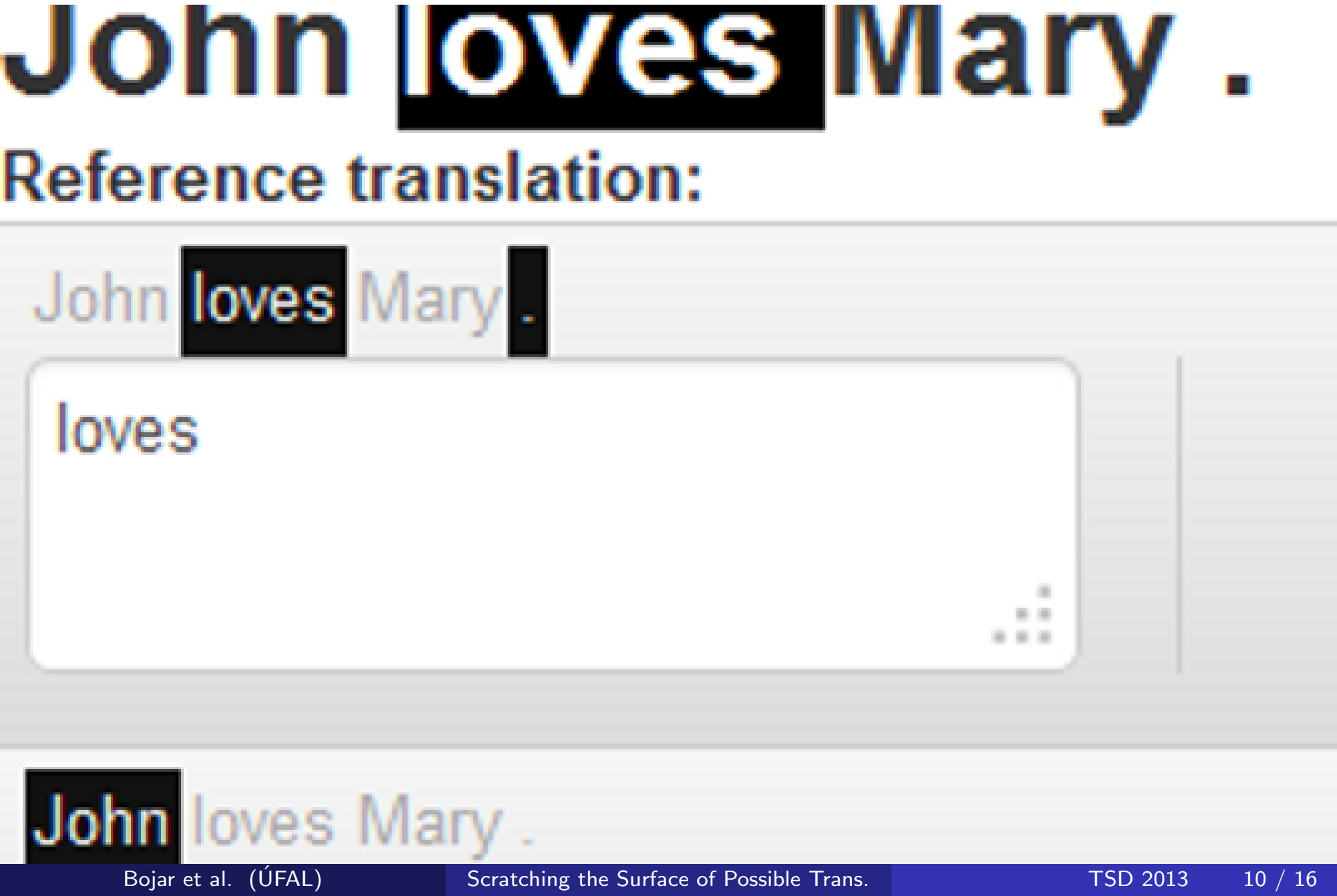


Web Interface

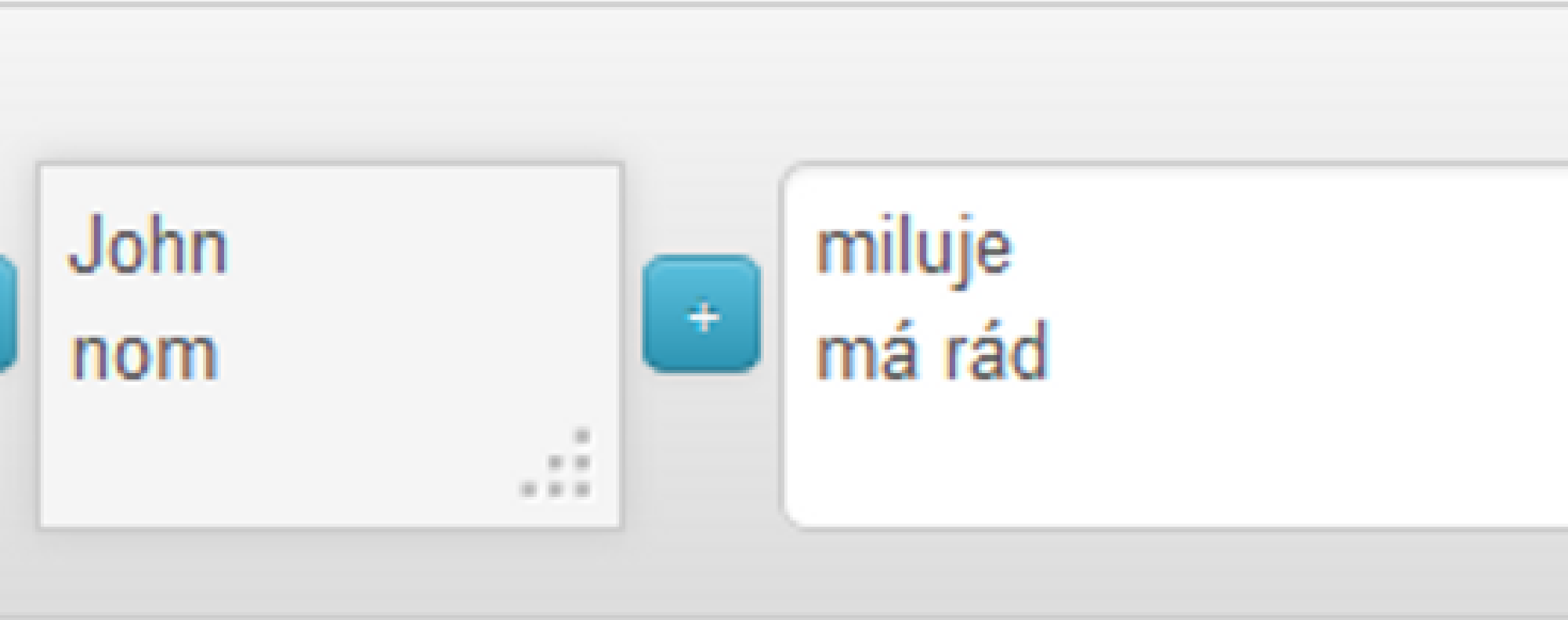
- Editing prolog source code is not for everyone.
- Our web-based graphical interface hides away the Prolog syntax:



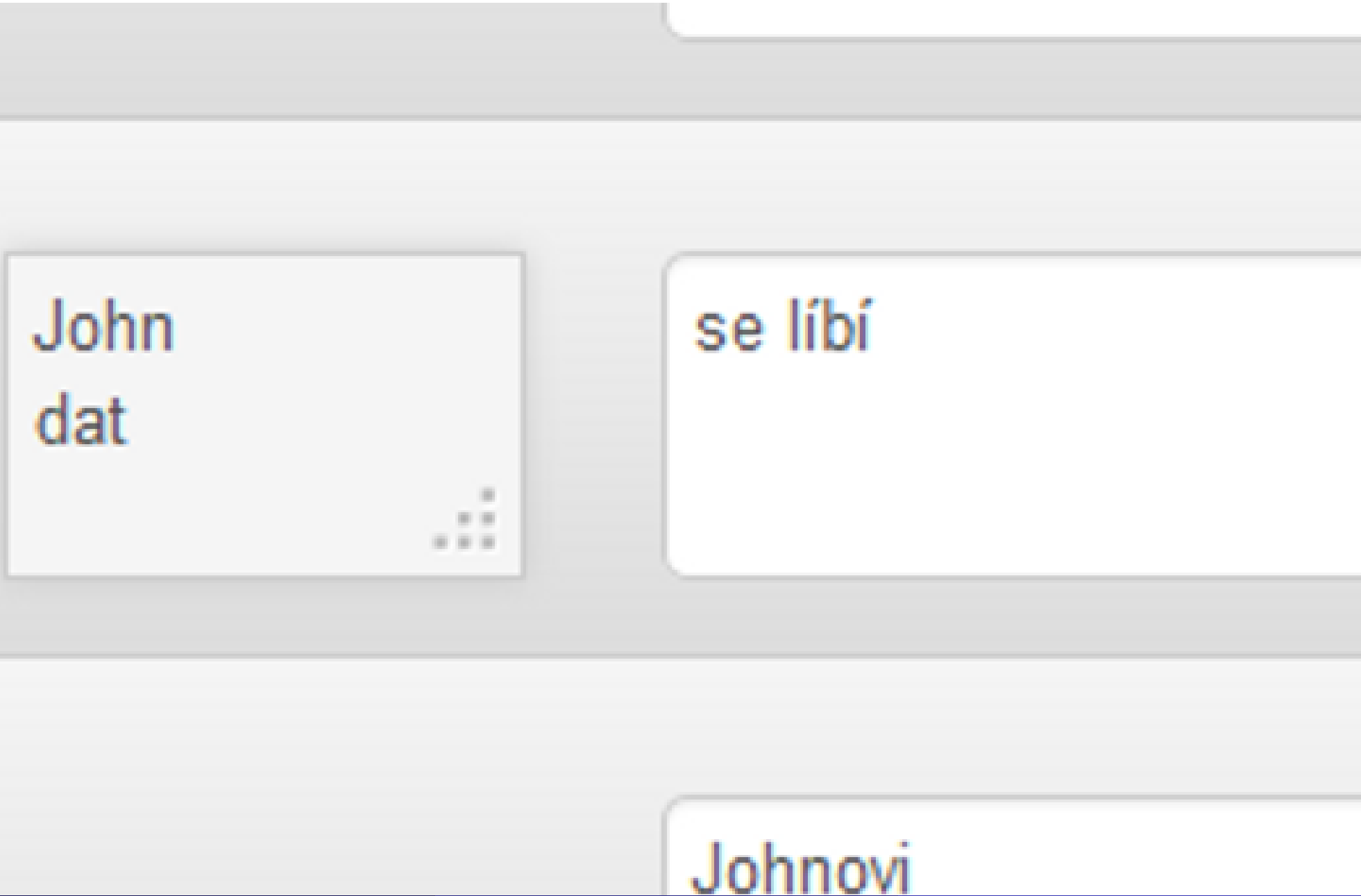
Web Interface



Web Interface



Web Interface



Collected Data

- We selected 50 sentences from the WMT11 test set for annotation

Collected Data

- We selected 50 sentences from the WMT11 test set for annotation
- 6 translators involved in the task, producing 77 sets of references
  - 24 sentences translated by one anotator
  - 25 sentences translated by two anotators
  - 1 sentence transalted by three anotators

Collected Data

- We selected 50 sentences from the WMT11 test set for annotation
- 6 translators involved in the task, producing 77 sets of references
  - 24 sentences translated by one anotator
  - 25 sentences translated by two anotators
  - 1 sentence transalted by three anotators

- Combined with other translations of WMT11 test set, we have:

Label	Reference set	Sent.	Refs.
O	The single official reference translation of WMT11	3003	1
P	Manual post-edits of a phrase-based MT system	1997	2
D	Our many references	50	Avg. 123k

Number of Obtained References

Annotation Interface	Avg. Sent. Length	Avg. Number of Refs.
Prolog	20.7	256k
Web	23.3	49k

Number of Obtained References

Annotation Interface	Avg. Sent. Length	Avg. Number of Refs.
Prolog	20.7	256k
Web	23.3	49k

- Prolog interface more efficient:
  - Much less screen space taken by each bubble.
  - Easier copying and modifications of bubbles.

Number of Obtained References

Annotation Interface	Avg. Sent. Length	Avg. Number of Refs.
Prolog	20.7	256k
Web	23.3	49k

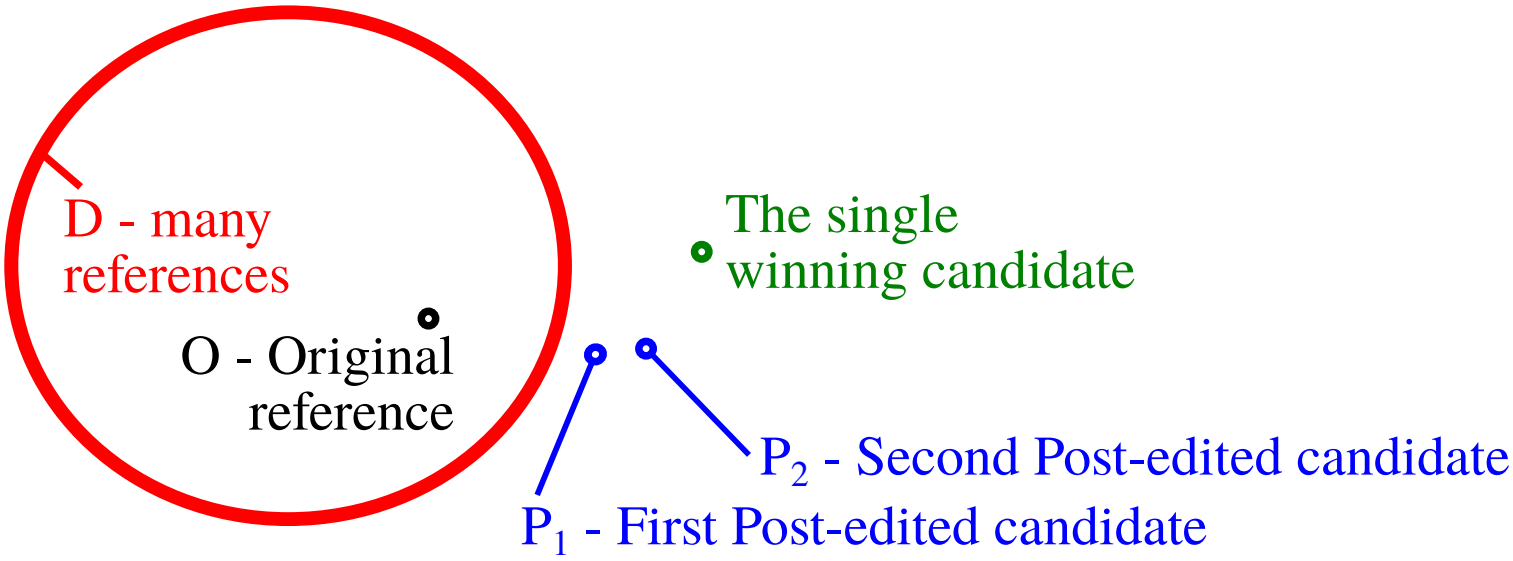
- Prolog interface more efficient:
  - Much less screen space taken by each bubble.
  - Easier copying and modifications of bubbles.
- About 5× more references per sentence were obtained in the 2 hours of annotation.

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations

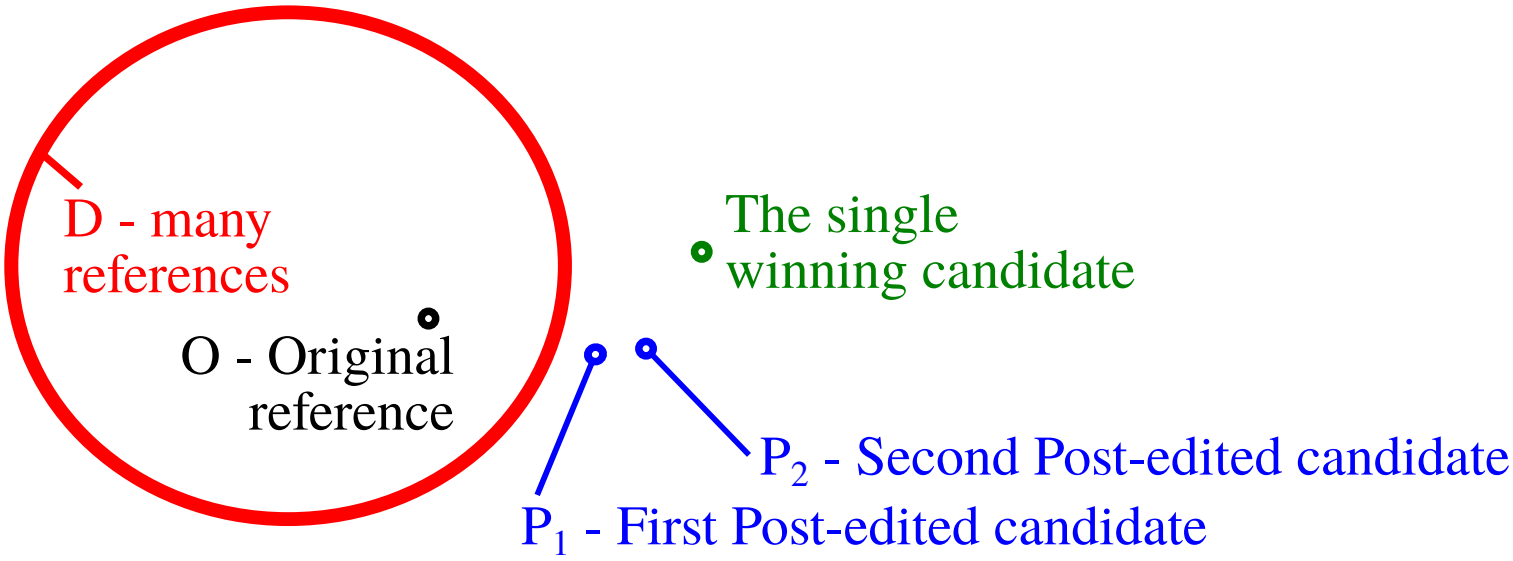
Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations



Similarity

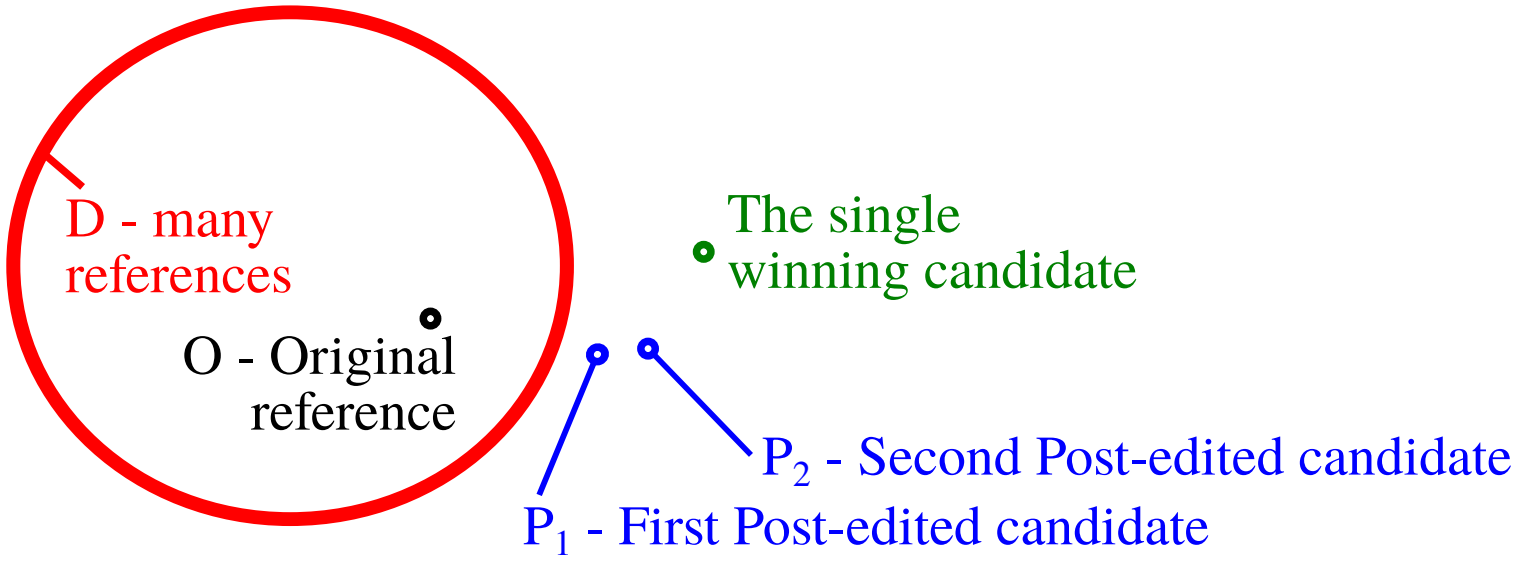
- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations



- We computed average similarity across all sentences and systems:

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations

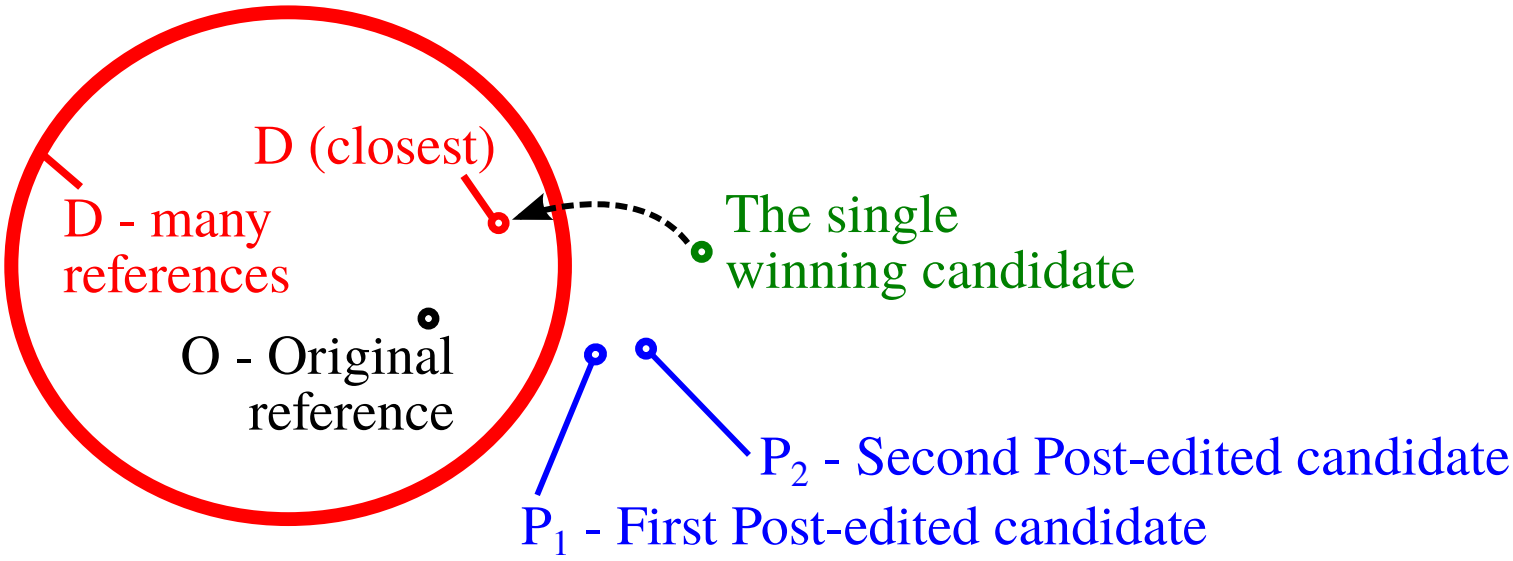


- We computed average similarity across all sentences and systems:

Reference Set Similarity

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations

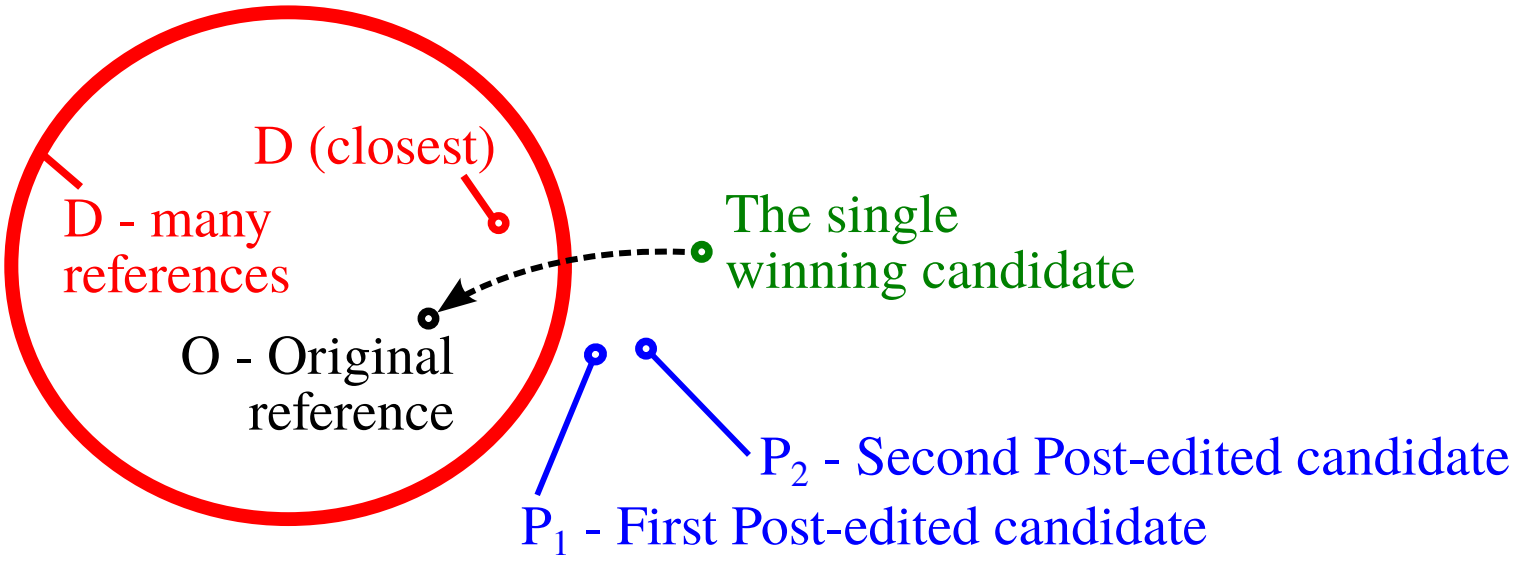


- We computed average similarity across all sentences and systems:

Reference Set	**D (closest)**
Similarity	0.62

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations

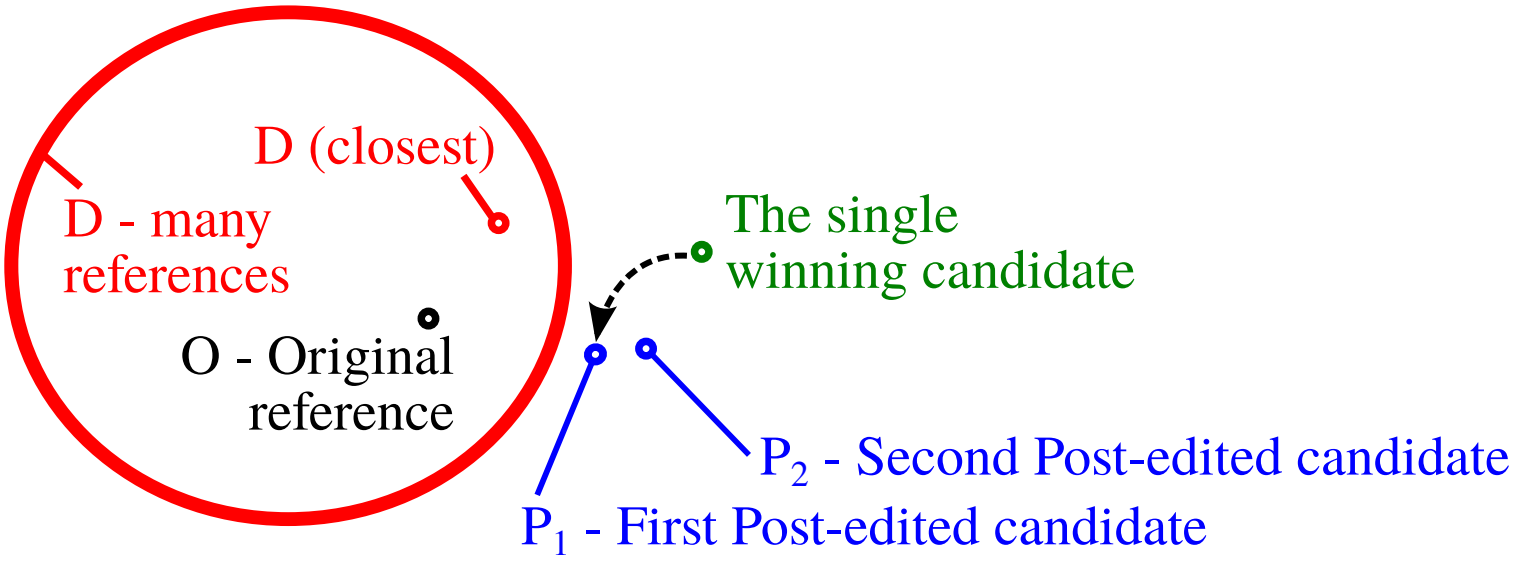


- We computed average similarity across all sentences and systems:

Reference Set	**D (closest)**	**O**
Similarity	0.62	0.51

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations

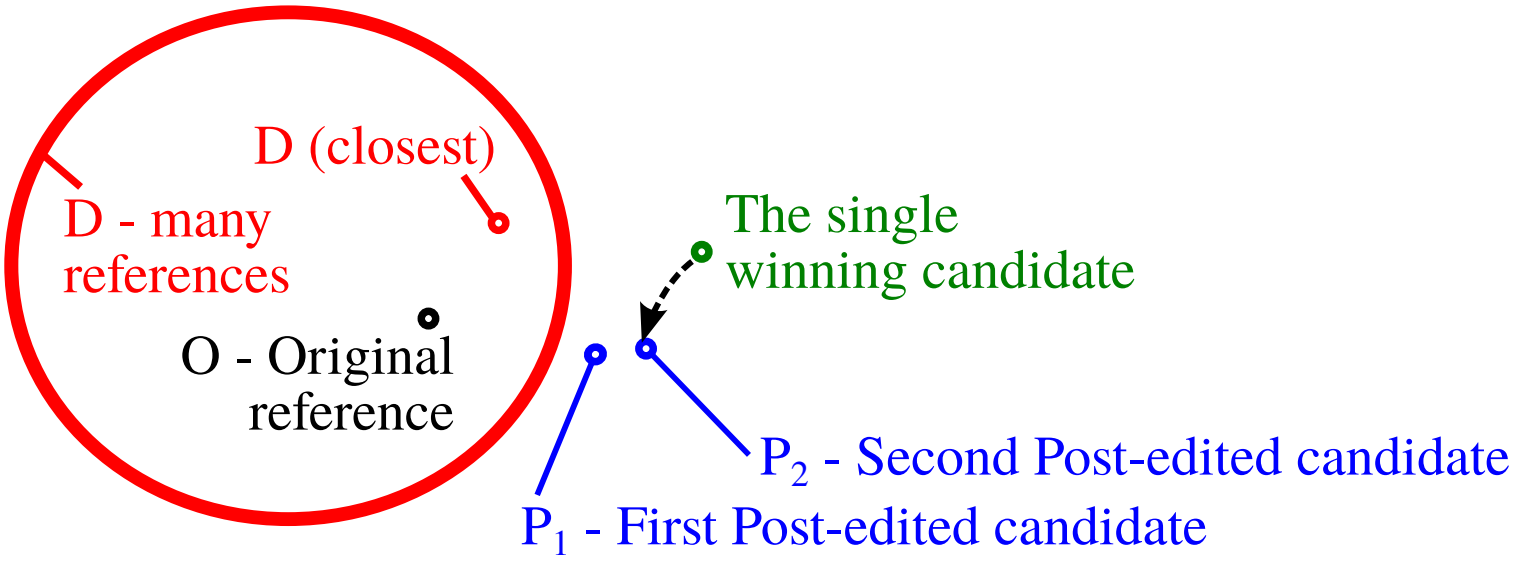


- We computed average similarity across all sentences and systems:

Reference Set	**D (closest)**	**O**	**P_1**
Similarity	0.62	0.51	0.63

Similarity

- For a given sentence and given MT system, we can measure similarity between candidate translation and one of the reference translations



- We computed average similarity across all sentences and systems:

Reference Set	**D (closest)**	**O**	**P_1**	**P_2**
Similarity	0.62	0.51	0.63	0.64

Example sentences

- **Src:** Great when a film has several target groups, but a shame if they are mutually exclusive.

Example sentences

- **Src:** Great when a film has several target groups, but a shame if they are mutually exclusive.
- **Orig Ref:** Je dobré, když má film více cílových skupin, je jen škoda, když se navzájem vylučují.

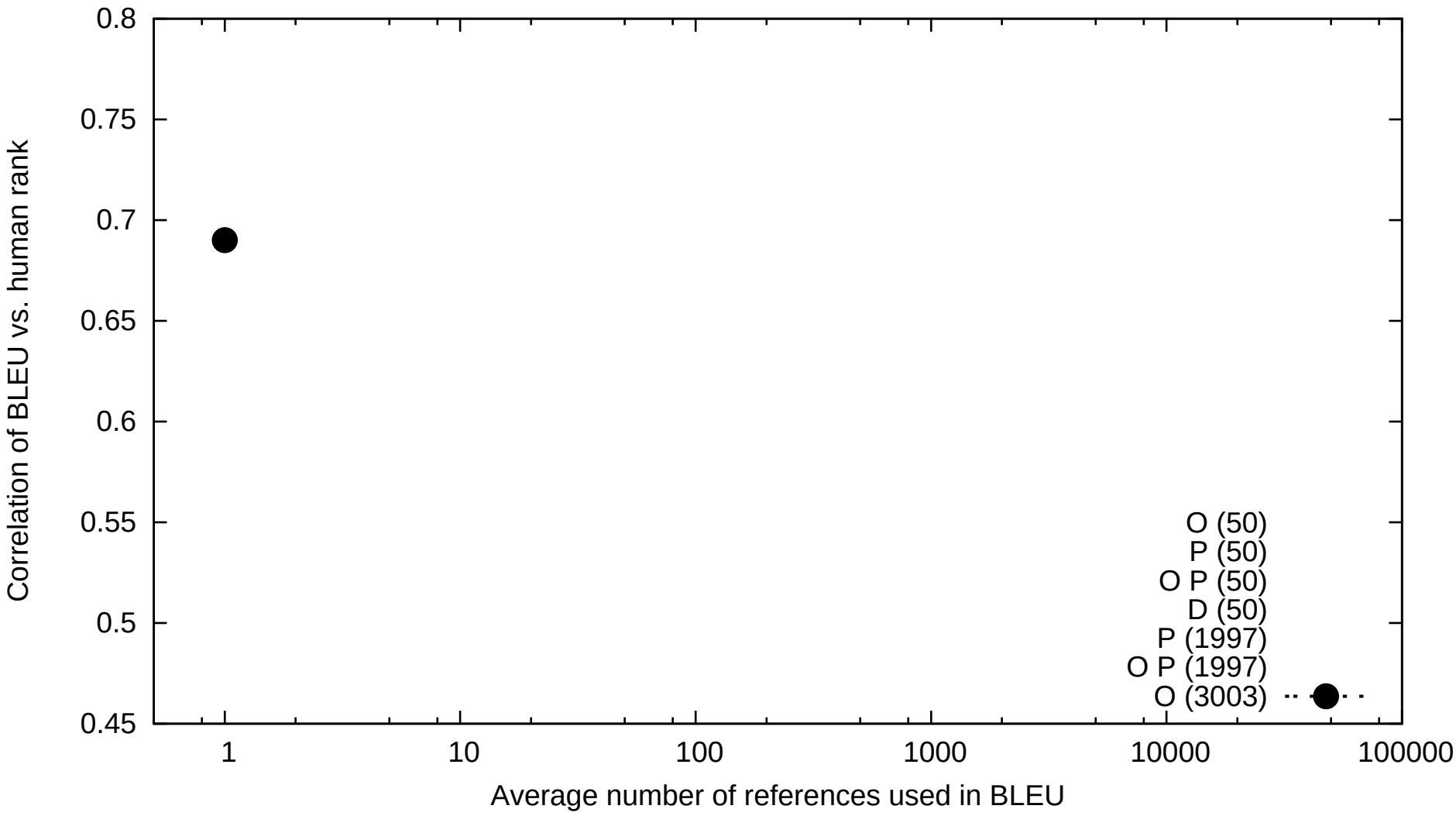
Example sentences

- **Src:** Great when a film has several target groups, but a shame if they are mutually exclusive.
- **Orig Ref:** Je dobré, když má film více cílových skupin, je jen škoda, když se navzájem vylučují.
- **Google:** Skvěle, když film má několik cílových skupin, ale škoda, kdyby se navzájem vylučují.

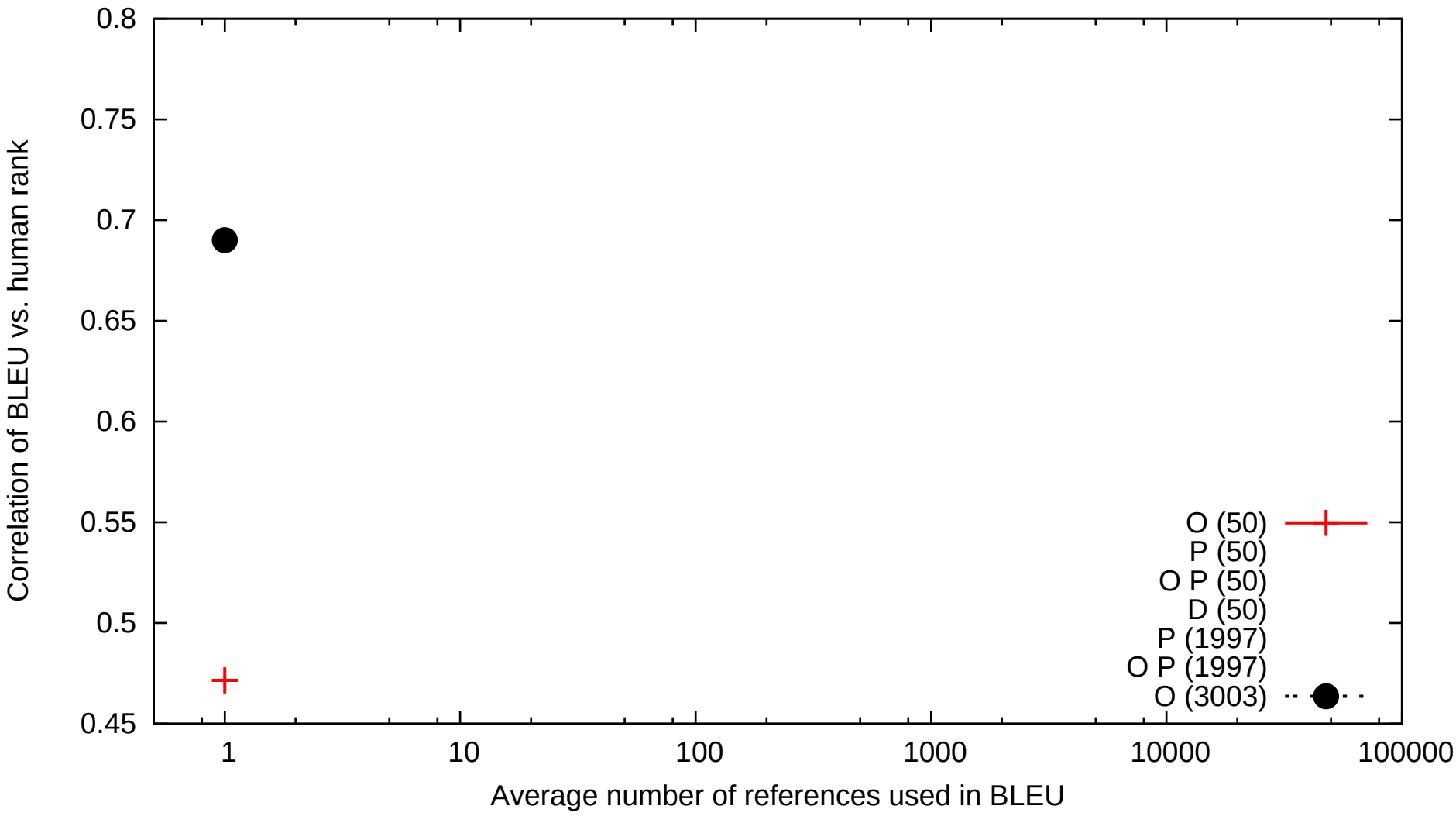
Example sentences

- **Src:** Great when a film has several target groups, but a shame if they are mutually exclusive.
- **Orig Ref:** Je dobré, když má film více cílových skupin, je jen škoda, když se navzájem vylučují.
- **Google:** Skvěle, když film má několik cílových skupin, ale škoda, kdyby se navzájem vylučují.
- **D(closest to Google):** Skvělé je, když film má několik cílových skupin, je ale škoda, když se tyto navzájem vylučují.
- **D(farthest from Google):** Výhodné je, pokud se snímek zaměřuje na víc cílových skupin, je jen smolné, pakliže jsou tyto skupiny vzájemně neslučitelné.

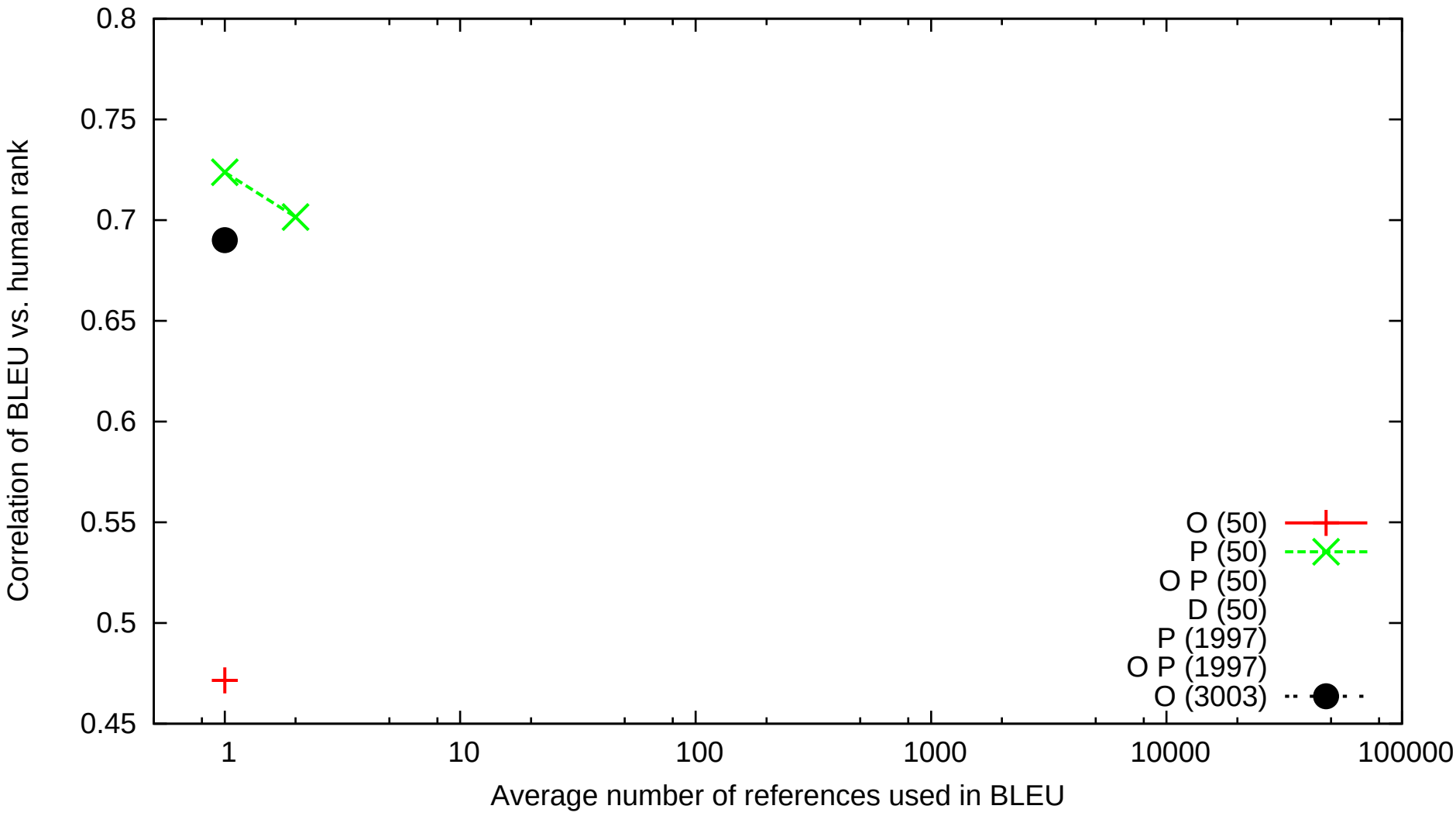
BLEU Correlations for Various Sets of References



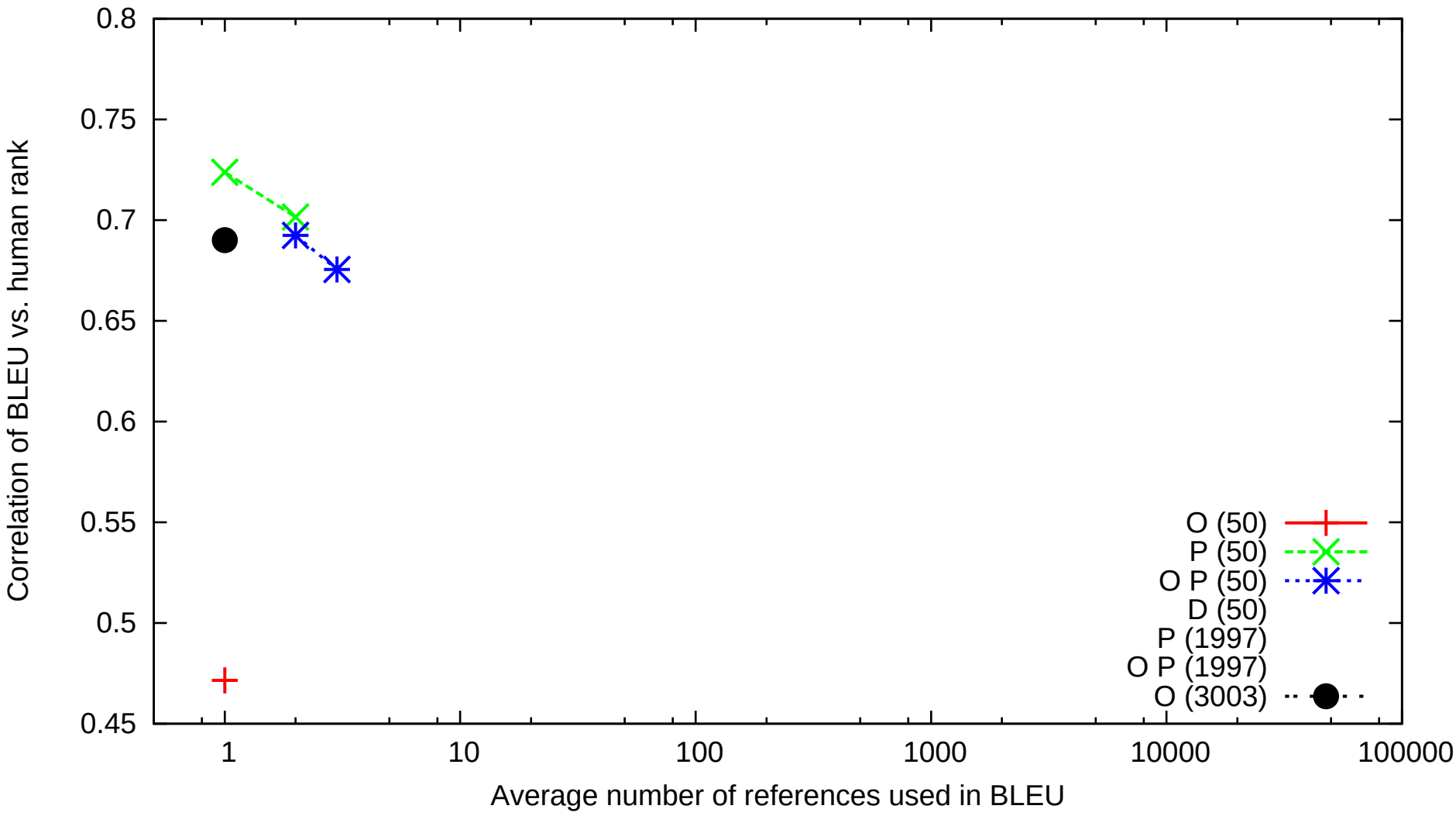
BLEU Correlations for Various Sets of References



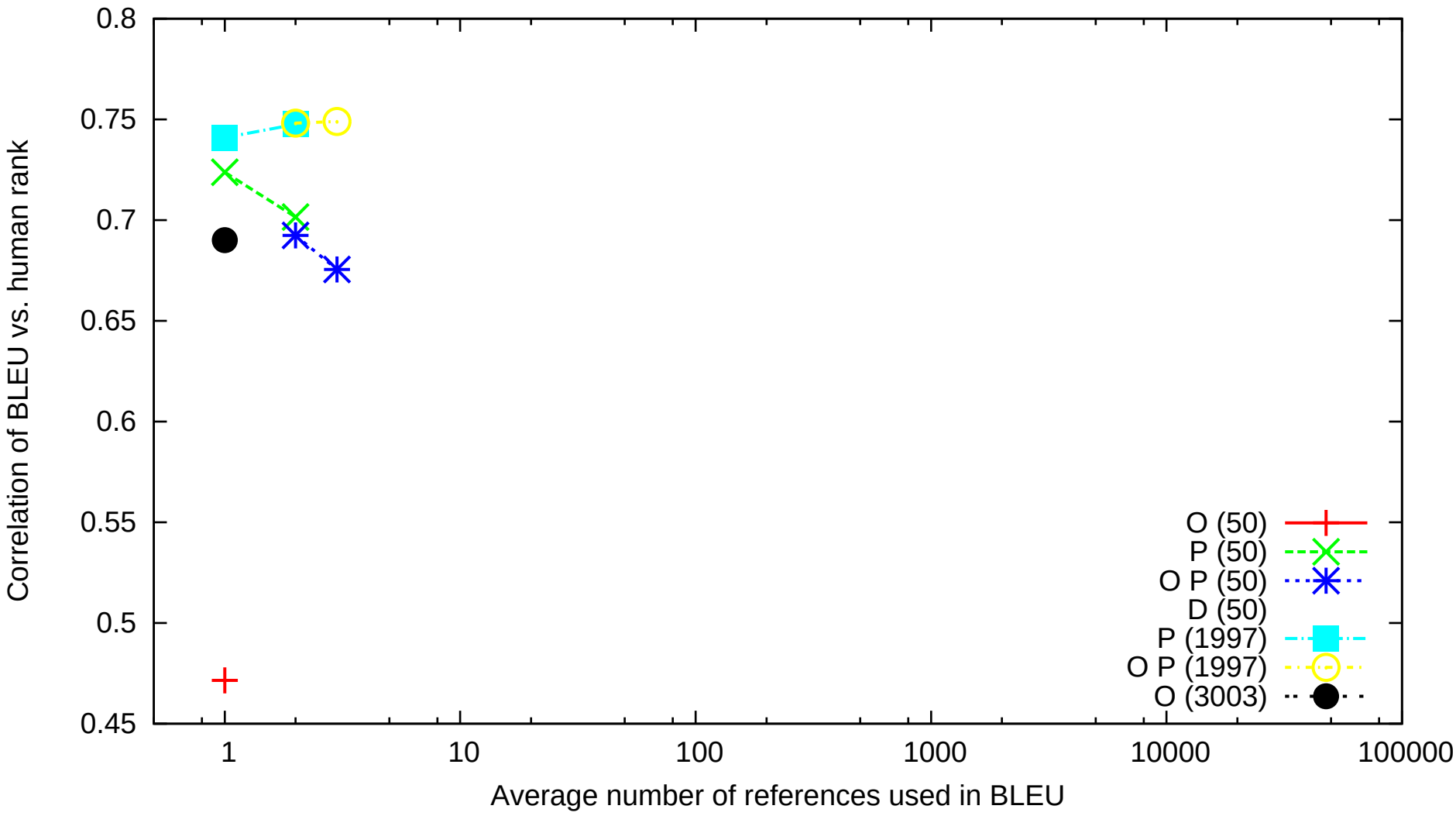
BLEU Correlations for Various Sets of References



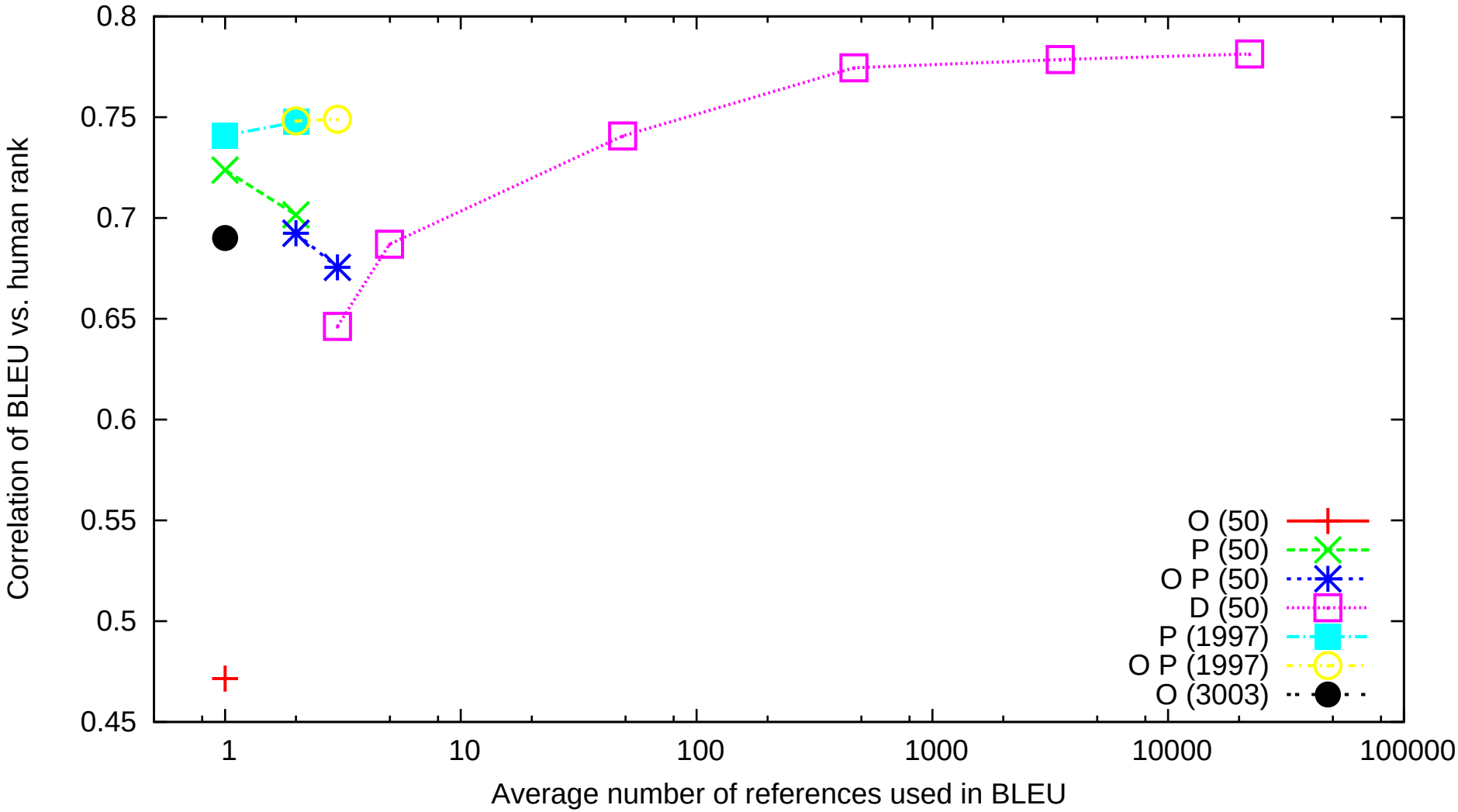
BLEU Correlations for Various Sets of References



BLEU Correlations for Various Sets of References



BLEU Correlations for Various Sets of References



Conclusion

- Proposed a unification-based compact representation for many refs.

Conclusion

- Proposed a unification-based compact representation for many refs.
  - Prolog interface more efficient than a web-based visual one.

Conclusion

- Proposed a unification-based compact representation for many refs.
  - Prolog interface more efficient than a web-based visual one.
- It is common to obtain dozens of thousand reference translations.

Conclusion

- Proposed a unification-based compact representation for many refs.
  - Prolog interface more efficient than a web-based visual one.
- It is common to obtain dozens of thousand reference translations.
- Some of these references are as close to the MT output as the post-edited version of the sentence.

Conclusion

- Proposed a unification-based compact representation for many refs.
  - Prolog interface more efficient than a web-based visual one.
- It is common to obtain dozens of thousand reference translations.
- Some of these references are as close to the MT output as the post-edited version of the sentence.
- Post-edited MT output serves as a better reference than regular translations.

Conclusion

- Proposed a unification-based compact representation for many refs.
  - Prolog interface more efficient than a web-based visual one.
- It is common to obtain dozens of thousand reference translations.
- Some of these references are as close to the MT output as the post-edited version of the sentence.
- Post-edited MT output serves as a better reference than regular translations.
- Our many references allow for a better correlation between manual and automatic MT evaluation techniques, despite the low number of sentences in the test set.