

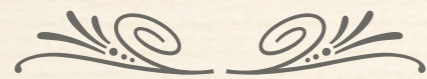
Not an Interlingua, But close:
Comparison of English AMRs to Chinese and Czech

—— Nianwen et al., 2014



Presented by Ziwei Peng
10/3/2015

What is AMR?



KEEP
CALM
AND
LOVE
AMR

Abstract Meaning Representation

- ❖ A rooted, directional and labeled graph
- ❖ Represents the meaning of a sentence
- ❖ Abstracts away from morph-syntactic idiosyncrasies, such as:
 - word category (verbs and nouns)
 - word order
 - function words (determiners, some prepositions)

Example

```
(b / be-located-at-91
  :ARG0 (p / person
         :name ( h / name
                 :op1 "Homer"
                 :op2 "Simpson")))
:ARG1 (a / amr-unknown)
:time (n / need-01
       :ARG0 (y / you)
       :ARG1 p))
```

Where is Homer Simpson when you need him?

However.....

- ❖ “AMR is heavily biased towards English. It is not an Interlingua.”

—— Abstract Meaning Representation for Sembanking (Laura.B et al., 2013)

- ❖ “AMR does not say anything about how it wants to be processed. It is closer to English than to other languages. It is not an interlingua.”

—— Abstract Meaning Representation (AMR) 1.2 Specification (Nathan.S et al., 2014)

Motivation

- ❖ refine the annotation standards for each of the three languages
- ❖ lead to more compatible annotation guidelines between the languages

Experiment

- ❖ Manually created AMRs for 100 Chinese and Czech sentences translated from English
 - 100 AMR-annotated sentences from a blog on Virginia road construction
 - Translated to Chinese and Czech and AMR-annotated in these two languages
 - English text has 1676 word and punctuation tokens
 - Annotated AMR representation contains 1231 nodes

a. Where is Homer Simpson when you need him?

(b / be-located-at-91

:ARG0 (p / person

:name (h / name

:op1 “Homer”

:op2 “Simpson”))

:ARG1 (a / amr-unknown)

:time (n / need-01

:ARG0 (y / you)

:ARG1 p))

English

c. Kde je Homer Simpson, když ho potřebujete?

(b / byt_umisten

:ARG0 (p / person

:name (h / name

:op1 “Homer”

:op2 “Simpson”))

:ARG1 (a / amr-unknown)

:time (p2 / potřebovat-1|need-01

:ARG0 (v / vy|you)

:ARG1 p))

Czech

b. 当你需要他时, 霍默 辛普森在哪里?

(b / be-located-at

:ARG0 (p / person

:name (n / name

:op1 ”霍默 辛普森|Homer Simpson”))

:ARG1 (a / amr-unknown)

:time (n2 / 需要|need

:ARG0 (y / 你|you)

:ARG1 p))

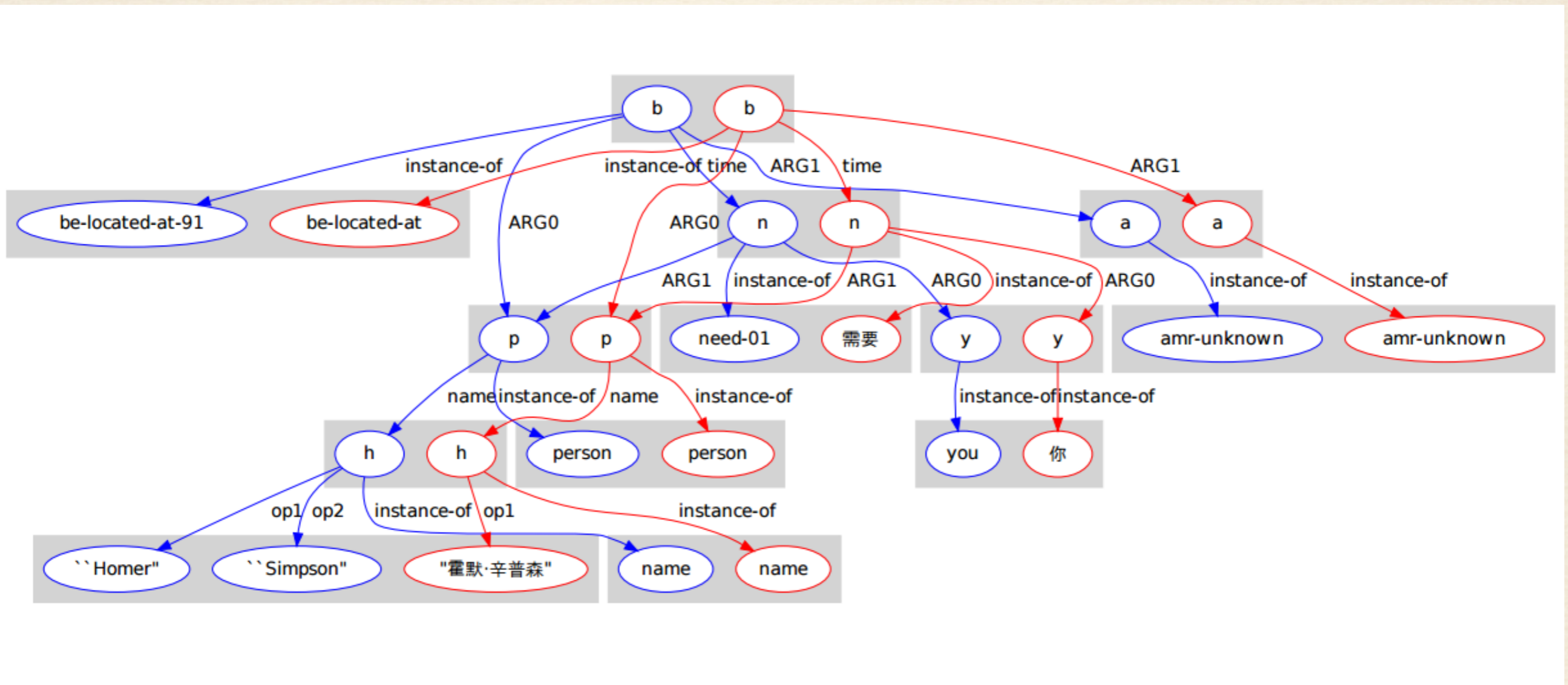
Chinese

English and Chinese AMRs

❖ Three scenarios

- 1. Translations of the same sentence are annotated with structurally compatible AMRs
- 2. Annotators of the different languages ended up with different AMRs, but the difference is a result of annotation choice and can potentially be reconciled
- 3. The differences in AMRs are due to different lexicalizations and such differences cannot be easily resolved without going to an even higher level of abstraction than the AMR representation currently provides

Structurally Compatible AMRs(Chinese/English)



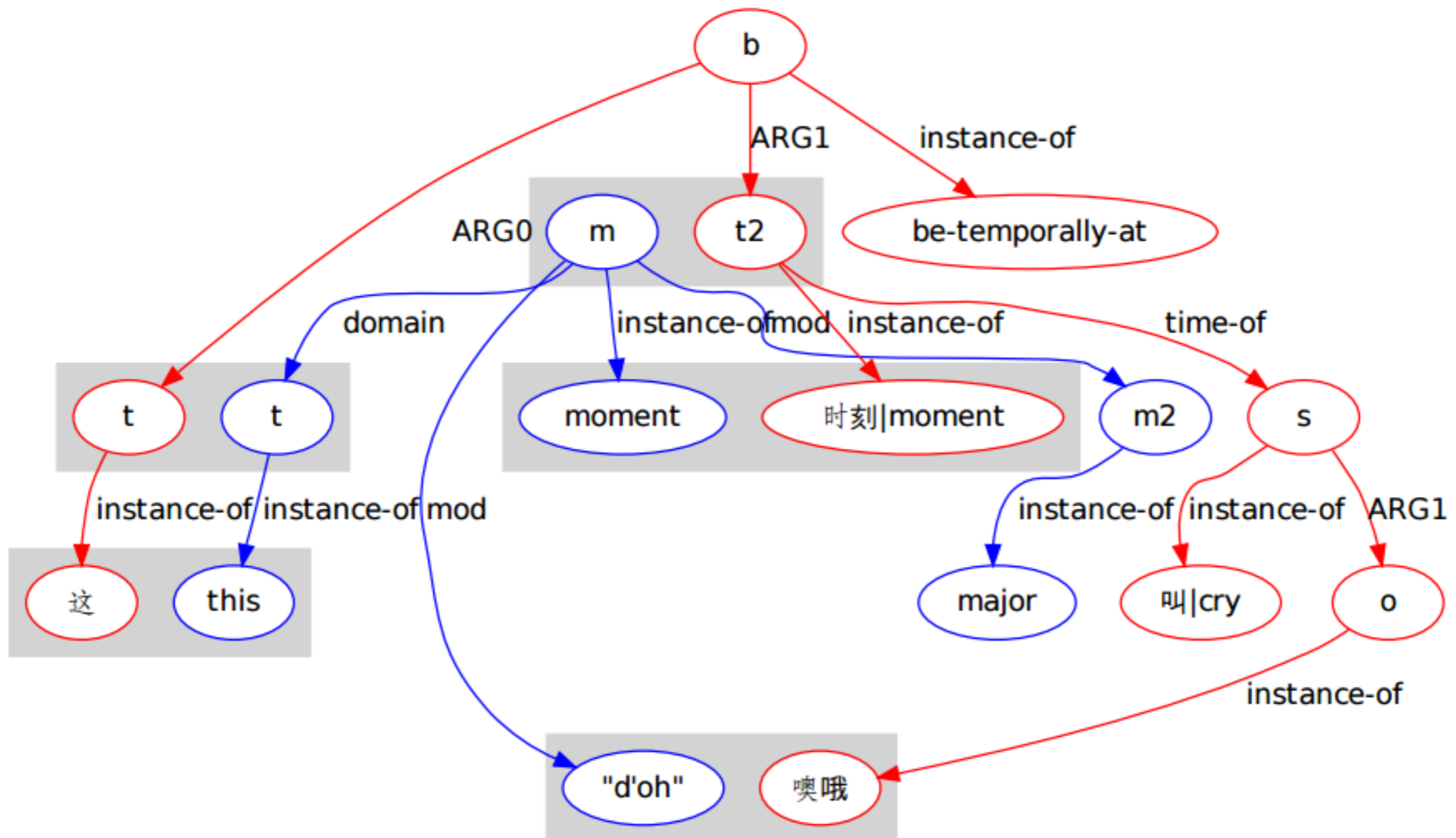
The AMRs of a Chinese sentence and its English translation show perfect alignment!

English and Chinese AMRs

❖ Three scenarios

- 1. Translations of the same sentence are annotated with structurally compatible AMRs
- 2. Annotators of the different languages ended up with different AMRs, but the difference is a result of annotation choice and can potentially be reconciled
- 3. The differences in AMRs are due to different lexicalizations and such differences cannot be easily resolved without going to an even higher level of abstraction than the AMR representation currently provides

Difference in annotation choice and lexicalization (Chinese/English)



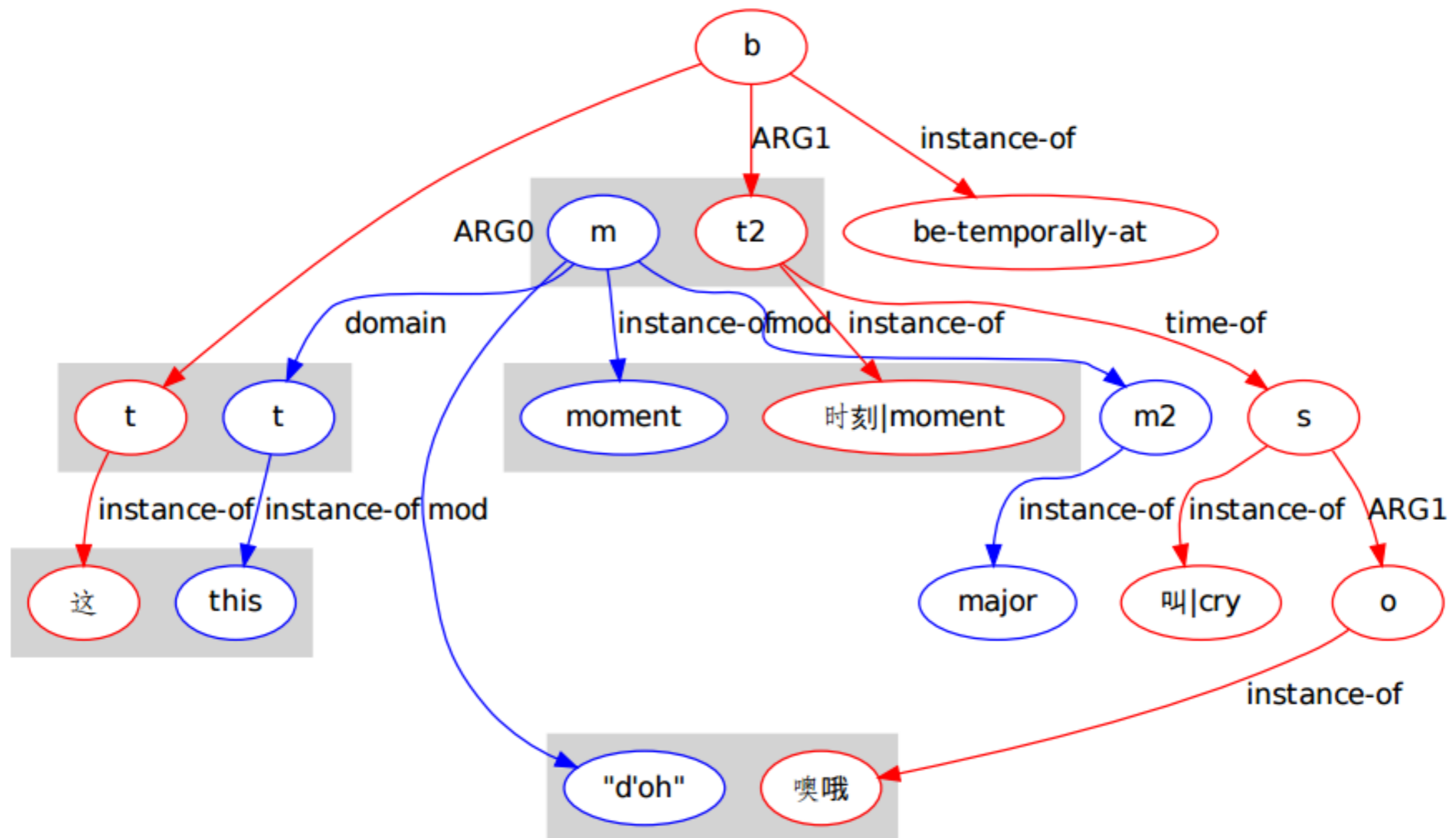
This is a major 'd'oh' moment.

English and Chinese AMRs

❖ Three scenarios

- 1. Translations of the same sentence are annotated with structurally compatible AMRs
- 2. Annotators of the different languages ended up with different AMRs, but the difference is a result of annotation choice and can potentially be reconciled
- 3. The differences in AMRs are due to different lexicalizations and such differences cannot be easily resolved without going to an even higher level of abstraction than the AMR representation currently provides

Difference in annotation choice and lexicalization (Chinese/English)

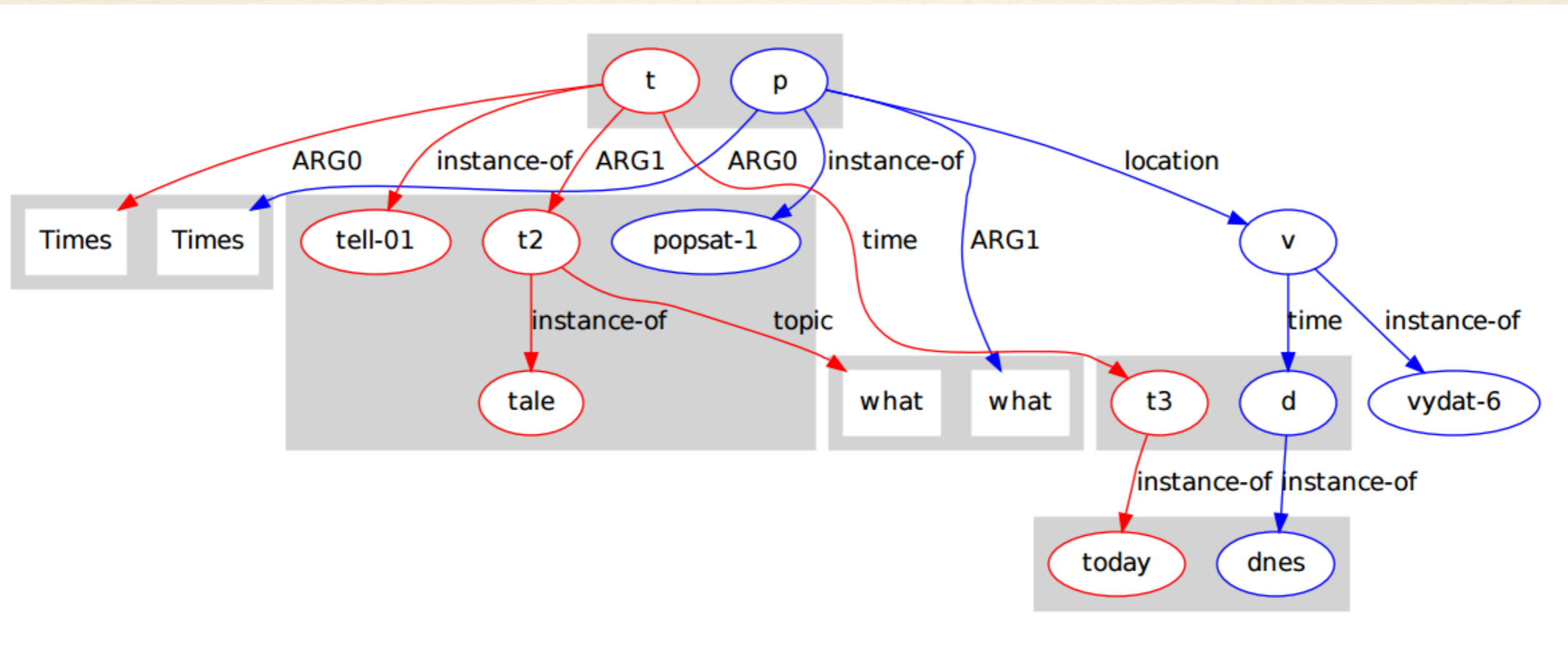


This is a major 'd'oh' moment.

English and Czech AMRs

- ❖ The AMRs for parallel sentences are either identical, differ in annotation choice, differ due to different ontologies, or are incompatible due to a substantial different sentence structure.

MWE and a structural divergence (Czech/English)



English and Czech AMRs

❖ Result:

- only 29 Czech sentences have structurally identical annotation
- 18 additional sentences contained differences which can be considered “local”

Conclusion

- ❖ The level of compatibility of AMR between English and Chinese is higher than between English and Czech
- ❖ A cross-linguistic comparison of English to Chinese and Czech AMRs reveals both cases where the AMRs for the language pairs align well structurally and cases of linguistic divergence
- ❖ Using AMRs as a sort of transfer layer in machine translation may require quite large and complex entries in the “translation dictionary”

Any questions?

Thank you !