



CLASSIFYING DIVERGENCES IN CROSS-LINGUAL AMR PAIRS

Shira Wein, Nathan Schneider

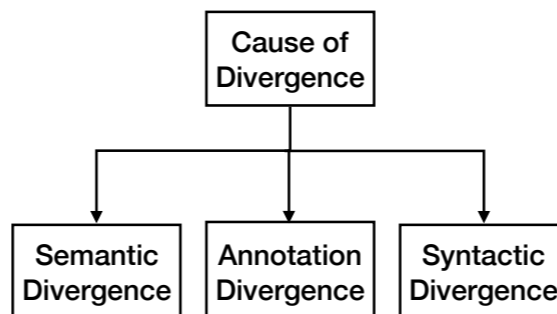
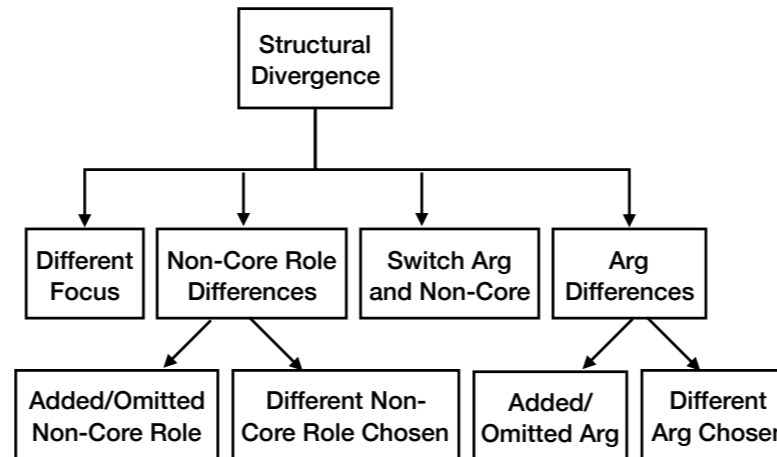
LAW-DMR 2021



MOTIVATION & APPROACH

- Parallel texts are often less than equivalent due to a range of translation divergences
- Identifying these translation divergences enables fuller utilization of parallel texts in downstream tasks (Vyas et al., 2018; Briakou and Carpuat, 2021)
- The Abstract Meaning Representation (Banarescu et al., 2013) formalism has been expanded to languages other than English in a variety of ways
- We propose an **annotation schema to identify specific points of divergence between parallel AMR graphs**
- Present a categorization schema to identify both type and cause of divergence, in order to:
 1. Make the data more adaptable to cross-lingual NLP applications
 2. Identify non-literal translations
 3. Make AMR more cross-lingual consistent (Song et al., 2019)
 4. Provide information on the way in which annotation, semantics, and syntax may play a role in cross-lingual AMR parsing (Damonte, 2019)
- Annotate a small corpus of English-Spanish parallel AMRs (Migueles-Abraira, 2017) from *The Little Prince* with our schema
- Analyze the ways in which cross-lingual parallel AMRs may differ in annotation

CLASSIFICATION SCHEMA



ANNOTATED EXAMPLES

focus.anno
 English: Volcanic eruptions are like fires in a chimney .
 Spanish: Las erupciones volcánicas son como el fuego de una chimenea .
 Literal translation: Volcanic eruptions are like the fire of a chimney.

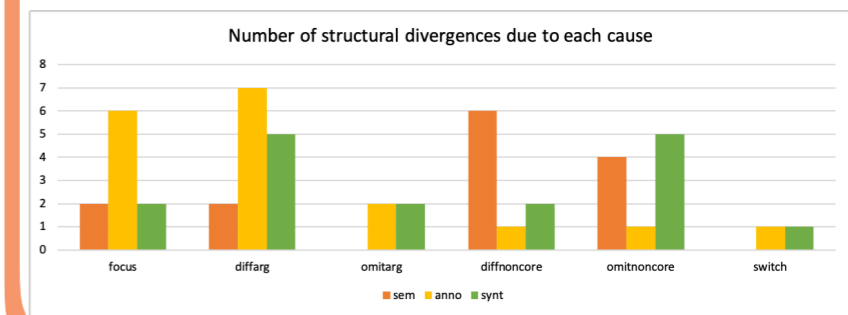
```

(e / erupt-0
:ARG1 (v / volcano)
:ARG1-of (r / resemble-01
:ARG2 (f / fire
:location (c / chimney)))
(p / parecer
:ARG0 (e / erupción
:mod (v / volcánico))
:ARG1 (s / ser-de-91
:ARG1 (f / fuego)
:ARG2 (c / chimenea)))
  
```

DISTRIBUTION OF CORPUS

	sem	anno	synt	
focus	2	6	2	10
diffarg	2	7	5	14
omitarg	0	2	2	4
diffnoncore	6	1	2	9
omitnoncore	4	1	5	10
switch	0	1	1	2
	14	18	17	

- Considerable difference in proportion of structural divergence due to each cause



- Limitations:** Currently only 50 Spanish AMR annotations publicly available
- Future Work:** Expand to additional language pairs for AMR or adapt to alternate meaning representation formalisms

Dataset: <https://github.com/shirawein/spanish-english-amr-corpus>

No divergence

English: Draw me a sheep ...
 (d / draw-01
 :ARG0 (y / you)
 :ARG1 (s / sheep)
 :ARG2 (i / i)
 :mode imperative)
 Spanish: Dibújame una oveja ...
 Literal translation: Draw me a sheep ...
 (d / dibujar
 :ARG0 (t / tú)
 :ARG1 (o / oveja)
 :ARG2 (y / yo)
 :modo imperativo)

diffarg.synt

English: The fourth planet belonged to a businessman .
 (b / belong-01
 :ARG0 (p / planet
 :ord (o / ordinal-entity :value 4))
 :ARG1 (b3 / businessman))
 Spanish: El cuarto planeta era de un hombre de negocios .
 Literal translation: The fourth planet was of a businessman.
 (p / pertenecer
 :ARG0 (p2 / planeta
 :ord (e / entidad-ordinal :valor 4))
 :ARG1 (h / hombre
 :mod (n / negocio)))

omitnoncore.synt (2)

English: That is funny!
 (f2 / funny
 :domain (t2 / that))
 Spanish: ¡ Qué gracioso !
 Literal translation: How funny!
 (g / gracioso
 :grado (t / tan))

focus.sem

English: Which is your planet?
 (p / planet
 :poss (y / you)
 :domain (a / amr-unknown))
 Spanish: ¿ De qué planeta eres ?
 Literal translation: What planet are you from?
 (s / ser-de-91
 :ARG1 (t / tú)
 :ARG2 (p / planeta
 :campo (a / amr-desconocido)))