A Corpus of German Abstract Meaning Representation (DeAMR)

Christoph Otto[1], Jonas Groschwitz[2], Alexander Koller[3], Xiulin Yang[4], Lucia Donatelli[5]
University of Potsdam[1], University of Amsterdam[2], Saarland University[3], Georgetown University[4], Vrije Universiteit Amsterdam[5]
christoph.otto@uni-potsdam.de[1], l.e.donatelli@vu.nl[5]

Abstract

We present the first comprehensive set of guidelines for German Abstract Meaning Representation (Deutsche AMR, DeAMR) along with an annotated corpus of 400 DeAMR. Taking English AMR (EnAMR) as our starting point, we propose significant adaptations to faithfully represent the structure and semantics of German, focusing particularly on verb frames, compound words, and modality. We validate our annotation through inter-annotator agreement and further evaluate our corpus with a comparison of structural divergences between EnAMR and DeAMR on parallel sentences, replicating previous work that finds both cases of cross-lingual structural alignment and cases of meaningful linguistic divergence. Finally, we fine-tune state-of-the-art multi-lingual and cross-lingual AMR parsers on our corpus and find that, while our small corpus is insufficient to produce quality output, there is a need to continue develop and evaluate against gold non-English AMR data.

Keywords: Abstract Meaning Representation, German, corpus creation, annotation, cross-lingual parsing

1. Introduction

Abstract Meaning Representations (AMRs) are semantic graphs that abstract away from surface syntax and capture the meaning of who does what to whom in a sentence (Banarescu et al., 2013). Though originally designed for English and decidedly not an interlingua, the AMR formalism exhibits certain properties that can be adapted to build AMR banks for other languages: for example, Spanish (Wein et al., 2022a), Chinese (Li et al., 2016), Vietnamese (Linh and Nguyen, 2019), Turkish (Azin and Eryiğit, 2019), Korean (Choe et al., 2020), Brazilian Portuguese (Anchiêta and Pardo, 2018) and Persian (Takhshid et al., 2022). Additional work on non-English AMR investigates *cross-lingual parsing*, where non-English text is parsed into English AMR (EnAMR) and evaluated against gold EnAMR (Damonte and Cohen, 2018). State-of-the-art cross-lingual parsers such as Uhrig et al. (2021) reach respectable performance on this task.

Here we present the first annotation guidelines and annotated corpus of German AMR (Deutsche AMR, DeAMR)[1]. Beyond providing this resource, we also investigate the questions: (i) how much does DeAMR differ from EnAMR? And (ii) is building language-specific AMR corpora worth it, when cross-lingual parsers exist? Our findings indicate that (i) there are meaningful differences, and (ii) yes, with some caveats.

In developing our guidelines, we find key linguistic differences between English and German relevant to AMR not easily remedied through translation. This is in line with related work finding that

```
(h / hinzufügen-02
  :ARG0 (s / sie)
  :ARG1 (z / Zeit
     :purpose (f / frühstücken-01)
     :time (b / bald))
  :time (d / darauf
     :degree (k / kurz)))
```

(a) “Es ist bald Zeit zum Frühstücken”, fügte sie kurz darauf hinzu.

```
(a / add-01
  :ARG0 (s / she)
  :ARG1 (t / think-01
     :ARG0 s
     :ARG1 (t2 / time
        :purpose (b / breakfast-01)))
  :time (a2 / after
     :quant (i / instant)))
```

(b) “I think it is time for breakfast,” she added an instant later.

Figure 1: PENMAN notation for parallel DeAMR and EnAMR sentences from *The Little Prince* (LPP).

AMR corpora in a range of languages do not perfectly align with EnAMR, notably due to structurally-rooted differences in how languages encode meaning (Xue et al., 2014; Urešová et al., 2014; Wein and Schneider, 2021). Qualitative analysis (Figure 1) and quantitative metrics corroborate these meaningful differences between DeAMR and EnAMR.

In this sense, our work is complementary to related work showing that in the task of obtaining

[1] https://github.com/chriott/DeAMR/

LREC-COLING 2024, pages 286–292
20-25 May, 2024.

EnAMR from sentences of different languages, source language has a dramatic effect on AMR structure (Wein et al., 2022b), whether the EnAMRs were obtained from parsers or human annotators. While this indicates that EnAMR has trouble capturing nuances of other languages, providing a strong argument for language-specific AMRs, we also find that training a standard model on our dataset of 400 sentences does not work well. More data and/or models better suited for low-resource settings may be required to have the benefits of language-specific AMRs outweigh the strengths of cross-lingual parsers.

We present complete annotation guidelines (Sec. 2) and a DeAMR corpus (Sec. 3). In a quantitative evaluation, we find inter-annotator scores comparable to previous AMR corpora, compare DeAMR and EnAMR structure, and evaluate both monolingual and cross-lingual parsers (Sec. 4). In Sec. 5, we make a case for the ongoing development of non-English AMR banks and highlight potential avenues for future work.

2. German AMR (DeAMR) Design

We take the guidelines for EnAMR[2] as our starting point to ensure that DeAMR can be integrated with ongoing AMR research and applications. Significant changes are made when EnAMR is inadequate to represent German linguistic phenomena.

Formatting. DeAMR follows several distinct conventions. i) Node and edge labels are case-sensitive and include all characters that are used in German, which can benefit German downstream parsing. ii) Coordinations and clausal connectives are migrated from AMR and remain as specified in the native guidelines. The exceptions are *or* and *and*, which we translate to the German equivalent *oder* and *und* to enhance human readability.

In agreement with several other non-EnAMR corpora (e.g Anchiêta and Pardo, 2018; Choe et al., 2020; Wein et al., 2022a; Li et al., 2016), we keep core and non-core role labels (e.g. `:ARGX`, `:location`, `:manner`), AMR-specific framesets (e.g. `be-located-at-91`, `have-polarity-91`) as well as canonical entity types (e.g. `government-organization`, `political-party`) in English.

German verb frames. The Universal PropBank (UP) project is used for German verb frame annotations (Jindal et al., 2022). UP is a multilingual semantic role labeling bank (including 23 languages from 8 language families) that builds on existing frame and role labels of the English PropBank by (Palmer et al., 2005). At this time, UP is still in development, which means that there is no full coverage of all existing predicate-argument frames that are present in the native English PropBank. Therefore, until the German PropBank (GPB) project covers a one-by-one range of frames, we agreed on a consensus to annotate missing information.

Annotations of German predicate senses which do not appear in the UP are made in one of two ways. The preferred choice is to find a nearly synonym German frame with a similar semantic meaning. For example, in the sentence *Ich bin nicht sicher ob...* ("I am not sure if..."), there is no suitable frame for 'sicher' ("to be sure"). Instead, we use `sicherstellen-01` ("to ensure"). If no suitable near-synonym exists, we create a new frame with `00`-sense numbering, using the roleset of the corresponding English frame for the argument structure (e.g. `responsible-03` has no matching German UP entry, therefore we use `verantworten-00` with the corresponding English argument structure). Generally, we abstain from creating framesets for other German word classes than verbs. This means in case an adjective evokes a frame that is available in German Propbank, we use it. Otherwise, we keep the non-verb word without any sense numbering.

Case marking for agreeing elements. German has a variety of morphological rules with respect to its case system that impact surface structure (Heinz and Matiasek, 1992). To ensure consistency in annotation, DeAMR uses a labeling convention for all agreeing elements including nouns, pronouns, adverbs, and adjectives. These elements are annotated in the AMR in nominative case with respect to the word they agree with in a sentence. For example, the adjective *lustig* ("funny") appears in the dative case in the LPP fragment *...von einer lustigen Stimme* ("from a funny voice"); we annotate the adjective as the nominative *lustige* in the AMR.

Formal pronouns. In German, the second person singular *du* ("you", informal) and third person plural *ihr* ("you all", informal) pronouns have a formal equivalent: *Sie* ("you" / "you all", formal). In DeAMR, formal pronouns are annotated with an additional `:polite +` label. This linguistic device enables to address a person of authority or is often used in a professional setting.

Compounds. Compounds are quite productive in German, such that new words are frequently invented. Here, the sparsity of coverage of German Propbank noted above can be problematic. To tackle this issue, we develop a heuristic to ensure consistent annotation of compounds in DeAMR (Figure 2). We first determine if the compound

[2] https://github.com/amrisi/amr-guidelines/blob/master/amr.md

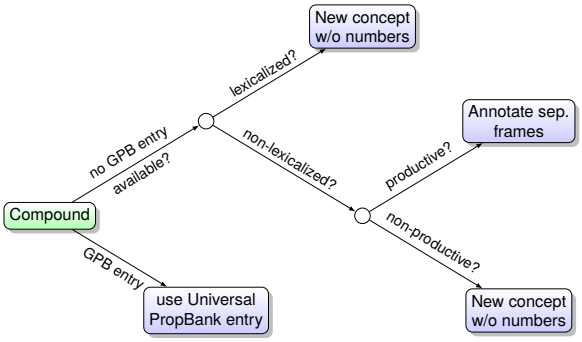

Figure 2: Decision tree for annotating compounds in DeAMR. Corresponding examples in text.

Genre/Text	Total Snt.	DeAMR	Avg. Tok.
Proxy (FT)	823	75	18.47
DFA (FT)	229	86	10.31
Bolt (FT)	133	33	8.5
Consensus (FT)	100	77	12.86
Xinhua (FT)	86	29	12.31
LLP	1562	100	10.4
Total	2933	400	12.92

Table 1: DeAMR corpus overview on genre, total sentences in the source corpora, subset of selected sentences for DeAMR annotation, and the average tokens per sentence in the DeAMR corpus.

has an entry in the German PropBank to utilize, e.g. *zufriedenstellen* `(z / zufriedenstellen-01)` ("to satisfy"), composed of the individual units *zu* "to", *Frieden* "peace", and *stellen* "to place". If not, we decide whether the compound is lexicalized, or if the separate word components are compositional and separable. If the compound is lexicalized, we create a new concept without verb sense numbering e.g. `(s / Schweinehund)` ("swine dog", but metaphorically used to refer to a person's lack of willpower). If the compound is compositional, we follow our decision tree and decide whether the compound is productive–i.e. the respective elements are frequently used in German compound formation. If productive, the individual components are annotated separately with the semantic head of the compound as the root e.g. *Vulkanausbruch* `(a / ausbrechen-02 :ARG1 (v / Vulkan))` ("volcanic eruption"). If not productive, we create a new concept without verb sense numbering e.g. `(h / Himbeere)` ("raspberry").

Modality. German exhibits more grammatical variation and nuance in modal semantics than English, about which readers are referred to the extensive literature (Kratzer, 2013; Portner, 2009). Our annotation nevertheless follows EnAMR practice: the meaning of modal markers, such as modal verbs, are mapped to a matching predicate frame that approximates the modal semantic meaning. Such annotation also serves to mark modal scope (c.f. Pustejovsky et al., 2019). DeAMR uses German PropBank frames for this task as specified in the full DeAMR guidelines. For instance, *Sie sollten kommen* ("They should come") translates to:

```
(e / empfehlen-01
    :ARG1 (k / kommen-01
             :ARG0 (s / sie)))
```

Modal particles (MP), frequent in German, are a subset of the particle word class and commonly used to help convey the speaker's emphasis or attitude towards a proposition (Bross, 2012); MPs usually impact the reading of the whole sentence in which they appear (Bross, 2012). While MPs can change their semantic meaning based on the mood of a sentence in which they appear, they do not change the proposition of a sentence and therefore have no influence on the truth-conditions, proving murky to pinpoint when designing a sentence-level semantic framework such as AMR (Van Gysel et al., 2021).

To handle MPs in DeAMR, we introduce a set of six new labels that leverage the existing `:mode`-role to indicate grammatical mood: `suprised`, `conclusive`, `confirmation-seeking`, `irony/sarcasm`, `confirming` and `resigning`. For example, depending on the context, the MP "ja" (closest English translation "still") in *Das hat ja funktioniert* ("This worked out") can emphasize the proposition that, despite low hope, something turned out well and can thus be marked with `:mode conclusive`. Additionaly examples and explanations are provided in the guidelines.

3. DeAMR Corpus & Annotation

Our goal with the DeAMR corpus was to cover a variety of different genres to represent a broad collection of German semantic phenomena. The sources for the DeAMR corpus are thus two different corpora (for which EnAMR annotations also exist): *The Little Prince* (LLP) (Antoine de Saint-Exupéry, 2015, 2018; Banarescu et al., 2013) and the *AMR 2.0 - Four Translation* dataset (FT) (Marco Damonte, Shay Cohen, 2020) (see Table 1). LLP contains a wide range of linguistic phenomena and has been annotated by other non-English AMR varieties, including Chinese (Li et al., 2016) and Spanish AMR (Wein et al., 2022a). FT is often used for cross-lingual parser evaluation (Wein and Schneider, 2022) and contains diverse genres of text, ranging from discussion forums, blog posts, transcript

```
(g / gehen-01 :mode imperative
    :ARG0 (d / du)
    :ARG3 (w / Weg))
```

(a) Mach dich auf den Weg!

```
(g / go-02 :mode imperative
      :ARG0 (y / you)
      :time (n / now))
```

(b) Now go!

Figure 3: Difference in EN-DE translation choice in *The Little Prince* (LPP).

of news broadcast, and newswire. We balanced sentence choice across the different domains of the FT corpus, (see Table 1), and manually select sentences that display interesting linguistic phenomena in German. The overall annotation took on average 9 minutes per DeAMR, which amounts to a total of 60 hours for 400 DeAMR. The corpus contains 400 sentences with an average token number of 12.92. The average token number of the LLP and FT source corpora together is 17.26.

4. Evaluation

Inter-Annotator Agreement & Analysis. We calculate inter-annotator agreement (IAA) between two AMR expert annotators to measure the effectiveness of the DeAMR guidelines and the quality of the corpus. We prepared a subset of 25 LLP and 15 FT sentences and report a Smatch score (Cai and Knight, 2013) of 0.7. Our result is promising given the current incompleteness of German PropBank and is comparable to other non-English IAA scores of 0.72 (BrP, Sobrevilla Cabezudo and Pardo, 2019), 0.79 (Korean, Choe et al., 2020) and 0.83 (Chinese, Li et al., 2016).

To better understand our annotation disagreements, we examined 50 annotation differences between Annotator 1 and Annotator 2. We found that of of those 50 differences, only 15 were German-specific, i.e. related to specifically German constructions (in contrast to English) or to the DeAMR guidelines. Among these 15, five differences were due to German PropBank being less complete than the English PropBank: when verb frames were missing, annotators had to choose from a range of workarounds, introducing ambiguities in the annotation process. The remaining 35 annotation differences resulted from general AMR disagreement inherent to EnAMR, too. This indicates that DeAMR annotation may be more difficult than English annotation, but not by much (note that it is also possible that there are English-specific difficulties to AMR annotation, that do not apply to DeAMR, which we cannot measure in this setup).

```
(a / anschauen-01
      :ARG0 (w / wir)
      :ARG1 (u / untergehen-01
              :ARG0 (s / Sonne)))
```

(a) Wir schauen uns den Sonnenuntergang an.

```
(l / look-01
      :ARG0 (w / we)
      :ARG1 (s / sunset))
```

(b) We look at the sunset.

Figure 4: Difference in annotation guidelines.

```
(a / add-01
  :ARG0 (s / she)
  :ARG1 (t / time
     :purpose (b / breakfast-01)
     :time (s2 / soon))
  :time (o / on
     :degree (s3 / short)))
```

"Es ist bald Zeit zum Frühstücken", fügte sie kurz darauf hinzu .

Figure 5: The translated DeAMR from Figure 1 following the methodology of Wein and Schneider (2021).

Divergences Between EnAMR & DeAMR. We classify causes and types of translation divergences between English and German cross-lingual AMR pairs to better understand AMR's utility as an interlingua. Following the methodology of Wein and Schneider (2021), we translate the node labels of 100 DeAMR (50 LPP, 50 FT) into English. We then evaluate the translated AMRs against the parallel gold EnAMRs with Smatch; we obtain an average score of 0.63, with variance from 0.12 to complete match of 1.0.

An example of a translated DeAMR (from Figure 1) that obtains a Smatch score of 0.55 can be seen in Figure 5. While this example is most closely aligned to the **focus,sem** structural divergence category of Wein and Schneider due to there being a different `:ARG1` in the translated version, the source of difference is the absence of the English hedge "*I think...*" in the German version. While this linguistic device exists in German, it is not as commonly used pragmatically; a similar story holds for the English translation in Figure 3. Such divergences point to potentially more nuanced divergences between languages that arise from lexical to pragmatic factors.

Regarding causes of structural divergence between EnAMR-DeAMR pairs, we similarly observe three main sources as Wein and Schneider. First, we note differences that stem from the translation choice (so-called *semantic divergences*); see

Figure 3. Second, we note differences that result because of annotator preference (*annotation divergences*), such as "keeping a journal", `(k / keep-03 :ARG1 (j / journal))`, where two German PropBank frames match the same English roleset `keep-03`, which reflects in separate DeAMR: `(f / führen-00 :ARG1 (t / Tagebuch))` and `(s / schreiben-01 :ARG1 (t / Tagebuch))`. Third, we note differences that stem from inherent differences between the languages (*syntactic divergences*). These differences are often codified in the DeAMR annotation guidelines themselves: new annotations for German-specific phenomena (i.e. UP, compounds, modality, pronouns) are established where there are meaningful differences in how the languages express meaning: for example (see Figure 4), "sunset"`(s / sunset)` and respectively "Sonnenaufgang" `(a / aufgehen-01 :ARG0 (s / Sonne))`. This kind of lexical divergence is not specifically outlined by Wein and Schneider in their hierarchy and points to the need for additional work on classifying cross-lingual divergences in AMR, especially divergences related to how features are assembled in the lexicon (e.g. Hoffman et al., 2013).

Parser Evaluation. We use several standard evaluation methods to assess the adequacy of our corpus. First, we evaluate the similarity between DeAMR and EnAMR gold pairs with XS2MATCH (Wein and Schneider, 2022). XS2MATCH is adapted from S2MATCH (Opitz et al., 2020) and evaluates the semantics shared by the AMR pairs in two languages based on LaBSE (Feng et al., 2022). This results in a high XS2Match F-Score of 0.82, indicating that DeAMR and EnAMR exhibit strong semantic similarity, though a gap still exists.

We additionally examine how a state-of-the-art cross-lingual AMR parser performs on our dataset. Following (Wein et al., 2022b)'s findings on Spanish AMR parsing, we adopt Uhrig et al.'s (2021) Translate-then-Parse (TP) approach. We run both the default fine-tuned T5 model and Bart-Large model. We also fine-tune a monolingual `T5-base` model for comparison (75% for training, 12.5% for dev, and 12.5% for test). We evaluate three settings: (1) XS2MATCH between the predicted AMRs and gold English AMRs; (2) SMATCH between predicted AMRs and gold German AMRs; (3) SMATCH score between predicted AMRs and the gold English AMRs.

Table 2 shows evaluation results from the TP approach. For TP, the results are quite different. The XS2MATCH scores of both models are quite high, which means they grasp the majority of the semantic meaning represented in AMR. However, the relatively low SMATCH scores within language (i.e., English) indicate that structural differences in equivalent AMRs persist. It is also worth noting that fine-tuning T5 on our corpus does not work: an F-Score of 0.275 with precision and recall being 0.431 and 0.202 on the test split is unsatisfactory and points to the need for larger datasets for non-English languages.

Evaluation	Precision	Recall	F-Score
TP(parse_xfm_bart_large)			
XS2MATCH(EN-DE)	0.808	0.769	0.788
SMATCH(EN-EN)	0.692	0.659	0.675
SMATCH(EN-DE)	0.420	0.434	0.427
TP(parse_t5)			
XS2MATCH(EN-DE)	0.806	0.760	0.782
SMATCH(EN-EN)	0.686	0.648	0.667
SMATCH(EN-DE)	0.419	0.428	0.423

Table 2: Results of TP on DeAMR Corpus.

5. Conclusion

We introduce the first annotated corpus of German AMR (Deutsche AMR, DeAMR) and comprehensive set of annotation guidelines. We additionally evaluate our corpus through inter-annotator measures and manual analysis of errors and translation divergence from English AMR (EnAMR). The significant adaptations we propose from EnAMR demonstrate EnAMR's inadequacy in providing complete semantic coverage of German linguistic phenomena, though significant overlap is noted. Our qualitative and quantitative evaluations further show evidence that subtle but meaningful structural differences exist across parallel cross-lingual data, especially at the level of the lexicon. Parsing evaluation confirms this. Future work will look closer at cross-lingual verb semantics and valency patterns as sources of translation and parsing divergence, given the predicative core design of AMR. We emphasize the need for continued development of non-English AMR corpora and resources to better understand the formalism and its limitations, as well as develop non-English-based evaluation methods.

6. Bibliographical References

Rafael Anchiêta and Thiago Pardo. 2018. Towards AMR-BR: A SemBank for Brazilian Portuguese language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Zahra Azin and Gülşen Eryiğit. 2019. Towards Turkish Abstract Meaning Representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 43–47, Florence, Italy. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Fabian Bross. 2012. German modal particles and the common ground. *Helikon. A Multidisciplinary Online Journal*, pages 182–209.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Hyonsu Choe, Jiyoon Han, Hyejin Park, Tae Hwan Oh, and Hansaem Kim. 2020. Building Korean Abstract Meaning Representation corpus. In *Proceedings of the Second International Workshop on Designing Meaning Representations*, pages 21–29, Barcelona Spain (online). Association for Computational Linguistics.

Marco Damonte and Shay B. Cohen. 2018. Cross-lingual Abstract Meaning Representation parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1146–1155, New Orleans, Louisiana. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

P. Gallmann and H. Sitta. 2007. *Deutsche Grammatik*. Lehrmittel der Interkantonalen Lehrmittelzentrale. Lehrmittelverl. des Kantons Zürich.

Wolfgang Heinz and Johannes Matiasek. 1992. *Argument structure and case assignment in German*. Österreichisches Forschungsinstitut für Artificial Intelligence.

Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods*, 45:718–730.

Angelika Kratzer. 2013. Modality for the 21st century. In *19th International Congress of Linguists*, pages 181–201.

Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 2016. Annotating the little prince with Chinese AMRs. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 7–15, Berlin, Germany. Association for Computational Linguistics.

Ha Linh and Huyen Nguyen. 2019. A case study on meaning representation for Vietnamese. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 148–153, Florence, Italy. Association for Computational Linguistics.

Juri Opitz, Letitia Parcalabescu, and Anette Frank. 2020. AMR similarity metrics from principles. *Transactions of the Association for Computational Linguistics*, 8:522–538.

Paul Portner. 2009. *Modality*. OUP Oxford.

James Pustejovsky, Ken Lai, and Nianwen Xue. 2019. Modeling quantification and scope in Abstract Meaning Representations. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 28–33, Florence, Italy. Association for Computational Linguistics.

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo. 2019. Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 236–244, Florence, Italy. Association for Computational Linguistics.

Reza Takhshid, Razieh Shojaei, Zahra Azin, and Mohammad Bahrani. 2022. Persian abstract meaning representation.

Sarah Uhrig, Yoalli Garcia, Juri Opitz, and Anette Frank. 2021. Translate, then parse! a strong baseline for cross-lingual AMR parsing. In *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, pages 58–64, Online. Association for Computational Linguistics.

Zdeňka Urešová, Jan Hajič, and Ondřej Bojar. 2014. Comparing Czech and English AMRs. In *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing*, pages 55–64, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

Jens EL Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, et al. 2021. Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3-4):343–360.

Shira Wein, Lucia Donatelli, Ethan Ricker, Calvin Engstrom, Alex Nelson, Leonie Harter, and Nathan Schneider. 2022a. Spanish Abstract Meaning Representation: Annotation of a general corpus. In *Northern European Journal of Language Technology, Volume 8*, Copenhagen, Denmark. Northern European Association of Language Technology.

Shira Wein, Wai Ching Leung, Yifu Mu, and Nathan Schneider. 2022b. Effect of source language on AMR structure. In *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, pages 97–102, Marseille, France. European Language Resources Association.

Shira Wein and Nathan Schneider. 2021. Classifying divergences in cross-lingual AMR pairs. In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 56–65, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Shira Wein and Nathan Schneider. 2022. Accounting for language effect in the evaluation of cross-lingual AMR parsers. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3824–3834, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Nianwen Xue, Ondřej Bojar, Jan Hajič, Martha Palmer, Zdeňka Urešová, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1765–1772, Reykjavik, Iceland. European Language Resources Association (ELRA).

7. Language Resource References

Antoine de Saint-Exupéry. 2015. *Der Kleine Prinz. Übertragen aus dem französischen Original von Alexander Varell.* Via tolino media.

Antoine de Saint-Exupéry. 2018. *The Little Prince. Translated by Irene Testot-Ferry*. Wordsworth Editions.

Banarescu, Laura and Bonial, Claire and Cai, Shu and Georgescu, Madalina and Griffitt, Kira and Hermjakob, Ulf and Knight, Kevin and Koehn, Philipp and Palmer, Martha and Schneider, Nathan. 2013. *Abstract Meaning Representation for Sembanking*. Association for Computational Linguistics.

Jindal, Ishan and Rademaker, Alexandre and Ulewicz, Michał and Linh, Ha and Nguyen, Huyen and Tran, Khoi-Nguyen and Zhu, Huaiyu and Li, Yunyao. 2022. *Universal Proposition Bank 2.0*. European Language Resources Association.

Marco Damonte, Shay Cohen. 2020. *Abstract Meaning Representation 2.0 - Four Translations LDC2020T07*. Linguistic Data Consortium, Philadelphia, PA, ISLRN 359-968-732-813-3.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.