Comparing Czech and English AMRs

Zdeňka Urešová   Jan Hajič   Ondřej Bojar
Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25, 11800 Prague 1, Czech Republic
{uresova,hajic,bojar}@ufal.mff.cuni.cz

Abstract

This paper describes in detail the differences between Czech and English annotation using the Abstract Meaning Representation scheme, which stresses the use of ontologies (and semantically-oriented verbal lexicons) and relations based on meaning or ontological content rather than semantics or syntax. The basic "slogan" of the AMR specification clearly states that AMR is not an interlingua, yet it is expected that many relations as well as structures constructed from these relations will be similar or even identical across languages. In our study, we have investigated 100 sentences in English and their translations into Czech, annotated manually by AMRs, with the goal to describe the differences and if possible, to classify them into two main categories: those which are merely convention differences and thus can be unified by changing such conventions in the AMR annotation guidelines, and those which are so deeply rooted in the language structure that the level of abstraction which is inherent in the current AMR scheme does not allow for such unification.

1 Introduction

In this paper, we follow on a previous first exploratory investigation of differences in AMR annotation among different languages (Xue et al., 2014), which has classified the similarities and differences into four categories: (a) no difference, (b) local difference only (such as multiword expressions vs. single word terms), (c) reconcilable difference due to AMR conventions, and (d) deep differences which cannot be unified in the AMR guidelines. In this paper, we would like to elaborate especially on the (b) and (c) types, which have been only exemplified in the previous work. In this paper, we would like to not only go deeper, but also present quantitative comparison on 100 parallel sentences, for all the aforementioned categories and some of their subtypes.

We will first describe the basic principles of AMR annotation (Banarescu et al., 2013) (Sect. 2, building also on (Xue et al., 2014)), then present the data (parallel texts) which we have used for this study (Sect. 3), and describe the quantitative and qualitative comparison between AMR annotation of English and Czech (Sect. 4). In Sect. 5, we will summarize and discuss further work.

2 Abstract Meaning Representation (AMR)

Syntactic treebanks in several languages (Marcus et al., 1993; Hajič et al., 2003; Xue et al., 2005) and related annotated corpora such as PropBank (Palmer et al., 2005), Nombank (Meyers et al., 2004), TimeBank (Pustejovsky et al., 2003), FactBank (Saurí and Pustejovsky, 2009), and the Penn Discourse TreeBank (Prasad et al., 2008), coupled with machine learning techniques, have been used in many NLP tasks. These annotated resources enabled substantial amounts of research in different areas of semantic analysis. There had already been tremendous progress in syntactic parsing (Collins, 1999; Charniak, 2000; Petrov and Klein, 2007) and now in Semantic Role Labeling because of the existence of the PropBank (Gildea and Jurafsky, 2002; Pradhan et al., 2004; Xue and Palmer, 2004; Bohnet et al., 2013)

and similar resources in other languages (Hajič et al., 2009), and TimeBank has fueled much research in the area of temporal analysis.

There have been efforts to create a unified representation which would cover at least a whole sentence, or even a continuous text (Hajič et al., 2003; Srikumar and Roth, 2013), and currently the Abstract Meaning Representation represents an attempt to provide a common ground for truly semantic and fully covering annotation representation.

An Abstract Meaning Representation is a rooted, directional and labeled graph that represents the meaning of a sentence and it abstracts away from such syntactic notions as word category (verbs and nouns), word order, morphological variation etc. Instead, it focuses on semantic *relations* between *concepts* and makes heavy use of predicate-argument structures as defined in PropBank (for English). As a result, the word order in the sentence is considered to be of little relevance to the meaning representation and is not necessarily maintained in the AMR. In addition, many function words (determiners, prepositions) that do not contribute to meaning and are not explicitly represented in AMR, except for the semantic relations they express. Readers are referred to Baranescu et al. (2013) for a complete description of AMR.[1]
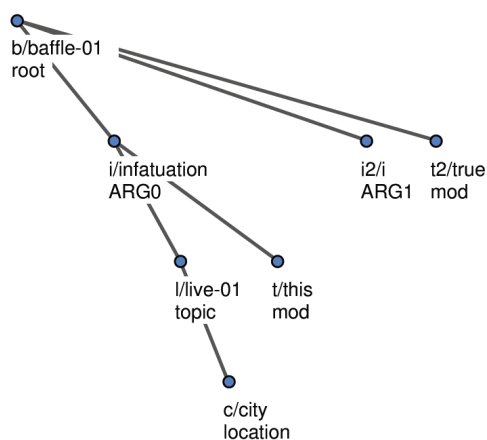

Figure 1: AMR annotation of the sentence *"This infatuation with city living truly baffles me."*

An example of an AMR-annotated sentence can be seen in Fig. 1. The predicate of the sentence (*baffle*) becomes the root of the annotation graph, with a reference to the correct sense `baffle-01` as found in PropBank frame files for `baffle`; PropBank frame files play the role of an ontology of events. Arguments of predicates, again as described in the PropBank frames, become the substitutes for roles of the "who did what to whom" interpretation - in the example sentence, *infatuation* - marked as `ARG0` - is the thing that baffles someone (the `ARG1`), i.e. *me* (the author of the text) in this case. This "baffling" is further modified by *"truly"*, and marked simply as a modifier, the semantics of which is fully represented by the word `true` itself. The agent (*infatuation*) has to be further restricted - it is the *"infatuation with city living"* which baffles the author - not just any infatuation. This is represented by the relation `topic` assigned to the edge between `infatuation` and `live-01` in the AMR graph, and the "living" (sense `live-01`) is further restricted by the `location` mentioned in the sentence, namely `city`. Finally, the modifier `this` is kept in, since it is needed for reference to previous text, where the "infatuation" has been first mentioned.

While the graphical representation in Fig. 1 is simplified in that it does not show the AMR's crucial `instance-of` relations explicitly as edges in the AMR graph, Fig. 2 shows the native underlying "bracketed" textual representation of the same tree, where the main nodes (i.e. those shown visibly in Fig. 1) are *mentions*, and the labels `baffle-01`, `true`, `live-01`, `city` etc. represent *links to external ontologies*. These links are currently represented only by these strings, or by links to PropBank

---

[1]This paragraph as well as the two preceding ones are taken over (and slightly adapted) from the introductory sections of a previous paper on this topic presented at LREC and co-authored by us (Xue et al., 2014); we share the same AMR formalism and data.

files for events. In the future, these links will be *wikified*, i.e. for concepts described in an external ontology, such as Wikipedia, they will be linked to it. The single- or two-letter "indexes" are in fact the labels (IDs) of the mentions, and they also serve for (co-)reference purposes; the slash ('/') is a shortcut for the `instance-of` relation.

```
(b / baffle-01
     :ARG0 (i / infatuation
           :topic (l / live-01
                 :location (c / city))
           :mod (t / this))
     :ARG1 (i2 / i)
     :mod (t2 / true))
```

Figure 2: Textual form of the AMR annotation of the sentence *"This infatuation with city living truly baffles me."*

In Czech, the event ontology has been approximated by the Czech valency dictionary, PDT-Vallex (Hajič et al., 2003), (Urešová, 2009), (Urešová and Pajas, 2009), (Urešová, 2006). No wikification of non-event nodes has been attempted yet; this is a continuing work, as it is for English.

3 The Data

We have drawn on a blog on Virginia road construction, taken from the WB part of the Penn Treebank. These sentences have already been annotated using AMRs, and also translated to Czech[2] and subsequently AMR-annotated. The English text has 1676 word and punctuation tokens (using the Penn Treebank style tokenization), and its annotated AMR representation contains 1231 nodes (not counting the `instance-of` nodes as separate nodes). The Czech version is a result of manually doubly-translated English original, which has been mutually checked and then one (slightly corrected) translation has been used for annotation. The Czech text has a total of 1563 tokens and its AMR representation contains 1215 nodes (again, not counting the `instance-of` nodes as separate nodes).

The data, once annotated, have been converted to a graph and in such a form presented to a linguist familiar with the AMR style annotation, to study and extract statistics for this comparison study. Fig. 3 shows such a side-by-side graphs for English and Czech AMR for the example parallel sentences.

4 The Comparison

4.1 Quantitative Comparison

In the first pass, we have concentrated on marking and counting the following phenomena:

- **structural identity**: sentences with identical structure have been marked as being structurally the same, even if some relation (edge) labels have been different

- **structural differences**: no. of structural differences have been noted in cases where one or more (sub)parts of the AMR graph differ between the two languages

- **local difference only**: out of the above, certain differences have been marked as "local only" - for example, if a multiword expression annotated as several nodes in one language corresponds to a single node in the other language

- **relation differences**: for each sentence, number of differences in relational labels has been counted

- **reference differences**: number of different references to an external ontology (or assumed differences in case no link to such an ontology was actually present in the annotation).

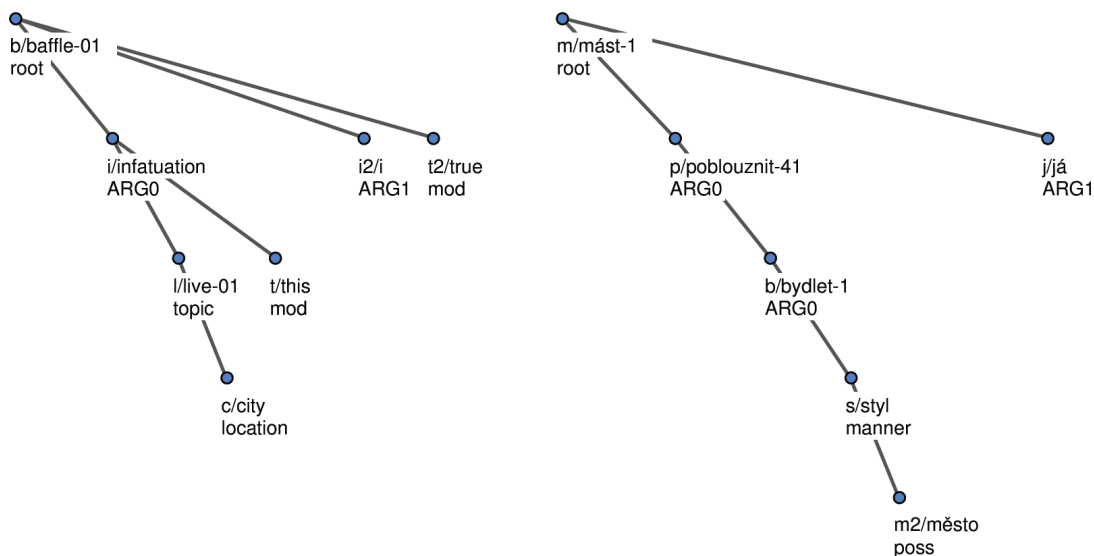[2]...and Chinese, but that was not used for this study.

Figure 3: AMR annotation of the sentence *"This infatuation with city living truly baffles me."* and its translation to Czech (*"Tohle/this poblouznění/infatuation bydlením/living-INSTR v/in městském/city-like stylu/style mě/me pořád/still mate/baffles."*)

It is obvious that we could have observed also other types of differences, but at this point, we wanted to have at least an idea how many differences exist in our approx. 1500-token sample. The resulting figures are summarized in Table 1.[3]

Same structure	Different substructures	Local difference only	Relation differences	Reference differences
29 (sents)	193 (subgraphs)	92 (subgraphs)	331 (nodes)	37 (nodes)
of 100	of approx. 800^2	of 193 (all diffs)	of 1215 Cz nodes	of 1215 Cz nodes
29 %	approx. 25 %^{2}	47.7 %	27.2 %	3.0 %

Table 1: Number and percentages of differences in the annotated data

The number of truly identically annotated sentences (including relation labeling) was only four, two of which has been interjective "sentences" at the beginning of the document ("Braaawk!"). On the other hand, 18 additional sentences would be structurally identical (on top of the 29) if local differences were disregarded, bringing the (unlabeled sentence identity) total to 47, or almost half of the data (47=29+18).

4.2 Analysis of Differences

The main goal of this study is to analyze differences in the annotation for the two languages, Czech and English, and determine if a reconciliation of the annotation is possible or not (and for what reason it is / it is not). Based on the above quantitative analysis, we have concentrated on relation labeling differences due their high proportion, and on structural differences due to their heterogeneous nature. The differences in reference annotation are small, but this is due to the lack of full referential annotation (it has been done for events, but only assumed for other types of entities due to the lack of ontology, or better to say due to the lack of "wikification" annotation in both languages), rather than due to high agreement. We will come back to this once wikification of the annotation is finished.

---

[3] Structural differences are hard to quantify exactly, since the base is difficult to define; it is part of future work.

4.3 Differences in Relation Labeling

The differences in function labeling should be taken with a grain of salt. The crucial question is what should count as a difference in relation labeling if the structure differs - should this be automatically counted as a difference, or not at all? In the figures summarized in Table 1, we have taken a middle ground: if the structural difference implied a change in labeling by itself, we have not counted that difference in order not to "penalize" the sentence annotation twice.

More detailed inspection of relation labeling differences, which appear to be relatively frequent at more than 1/4th of all nodes in the annotation, revealed that the by far most frequent mismatch is caused by different argument labeling for events.[4] While for most purely transitive verbs there is a complete match, for most other there is a discrepancy due to the attempted semanticization of PDT-Vallex argument labels ADDR (addressee), EFF (effect) and ORIG (origin), while PropBank simply continues to number arguments of corresponding verbs consecutively (for example, *I thought there is.ARG1 ...* vs. *Myslel/I-thought jsem, že/that tam/there je.ARG3←EFF/is ...*). The concept of "shifting" in PDT-Vallex, which compulsorily fills the first two arguments on syntactic grounds as ACT(ARG0) and PAT(ARG1) is another source of differences. Furthermore, PropBank leaves out ARG0 e.g. for unaccusative verbs (for example *The window.ARG1 broke* vs. *Okno.ARG0←ACT/window se/itself rozbilo/broke.*). Finally, some differences are due to some arguments not being considered arguments at all in the other language, in which case some other AMR label is used instead (for example, *We could have spent 400M.ARG3 ... elsewhere* vs. *... mohli/could utratit/spend 400M.extent ... jinde/elsewhere*).

These differences could possibly be consolidated (only) by carefully linking the two lexicons (with AMR guidelines intact). This is in fact being performed in another project (Sindlerova et al., 2014), but it is a daunting manual task, since the underlying theories behind PropBank and PDT-Vallex/EngVallex differ. However, one has to ask if it does make sense to do so, because with enough parallel data available, the mappings can be learned relatively easily: in most cases, no structural differences are involved and there will be a simple one-to-one mapping between the labels (conditioned on the particular verb sense).

4.4 Structural Differences

Local differences can be safely ignored, since they will be in most cases resolved during the assumed process of wikification, i.e., linking to an ontology concept. For example, the abbreviation *VDOT* (Virginia Department of Transportation), which has to be (and was) translated into Czech in an explanatory way (otherwise the sentence would become not quite understandable, if only because of the real-world context). Without wikification, it could not be linked as a whole, and thus a subgraph has been created with the AMR-appropriate internal semantic relations in the translation (e.g. *Virginia.location*, etc.).

Certain differences, albeit "localized" into a small subtree (or subgraph) corresponding to a single node or another small subtree (subgraph), cannot be resolved by wikification of a different event ontology (than PropBank or PDT-Vallex). For example, light verb constructions or even certain modal or aspectual constructions could have a single verb equivalent resulting in two node vs. single node annotation: *get close* vs. *přiblížit-se*, *make worse* vs. *zhoršit*, *take position (for sb)* vs. *zastávat-se* or *causing sprawl* vs. *roztahuje-se*.

Looking at the true structural differences, we have found that there are actually quite a few reasons for them to appear in the annotation. We will describe them in more detail below.

Non-literal translation is the primary reason for such differences.[5] For example, *destination* vs. *kam/where-to lidé/people jezdí/drive* (Fig. 4), or *job center* vs. *místo/place, kde/where pracuje/work hodně/many lidí/people*; these cases cannot be unified neither by changing the translation to a more literal one (because it would be strongly misleading in the given context, despite the fact that literal translation of both *destination* as well as *job center* does exist in Czech), neither by changing the guidelines, since the level of abstraction of AMR does not call for a unification of such concepts. Sometimes, non-literal

---

[4] The Czech PDT-Vallex argument labels have been mapped to PropBank labels as follows: ACT → ARG0, PAT → ARG1, ADDR → ARG2, EFF → ARG3 and ORIG → ARG4.

[5] This includes also cases of truly wrong translation, stemming of translator's misunderstanding of the facts behind the sentence. This has been found fairly often only after we studied the differences in depth, since a superficial reading and standard translation revision procedure did not help.

translation is forced upon the translation because no word-for-word translation exists, such as in *in the aggregate*, which has to be translated using an extra clause *z/from celkového/overall pohledu/view to/it je/is tak/so, že/that ...* (Fig. 5).
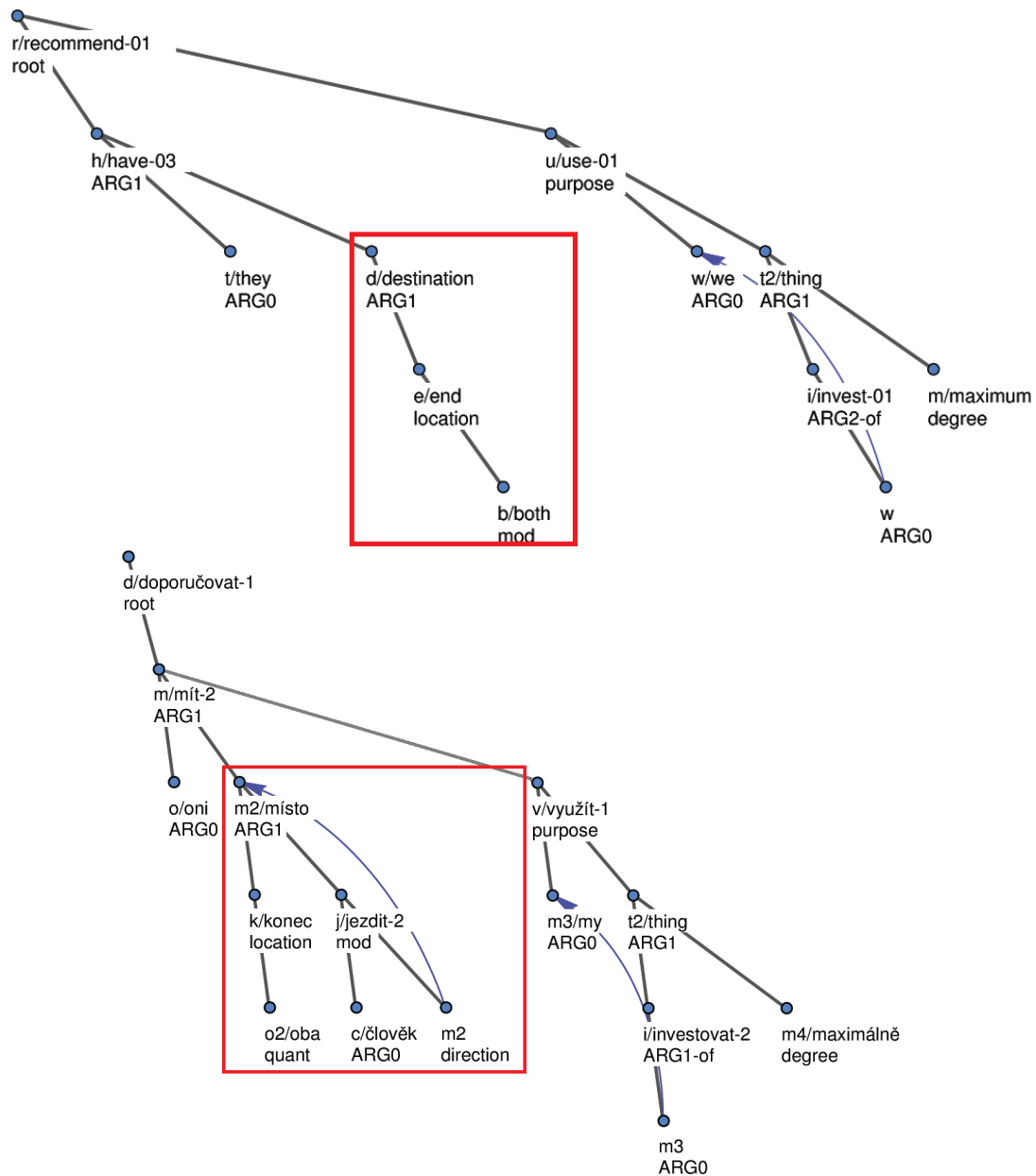

Figure 4: AMR structural difference: *destination* vs. *kam/where-to lidé/people jezdí/drive*

Phraseological differences and idioms form another large group of differences between the two languages. The possibility of changing the translation is even more remote than in the above case, even if we had the chance: the provided translation is actually the correct and perfect one. The reason for different annotation lies in the AMR scheme, which does not go that far to require "unified" annotation in such cases where the idiom or specific phrase cannot be linked to the external ontology as a single unit. For example, English "I don't see any point" is translated as "nemá/not-have smysl/purpose", and despite the fact that `have-purpose-91` is a specific event reference in English (and has been used in the annotation), the verb "see" still remains annotated as a separate event node, which is not the case in Czech, since no "seeing" is expressed in the sentence and it could hardly be asked for in the guidelines to be inserted. Similarly, *I commute back and forth* has been translated simply as *dojíždím/commute*, which is semantically perfectly equivalent but the *back and forth* has been kept in the English annotation, because
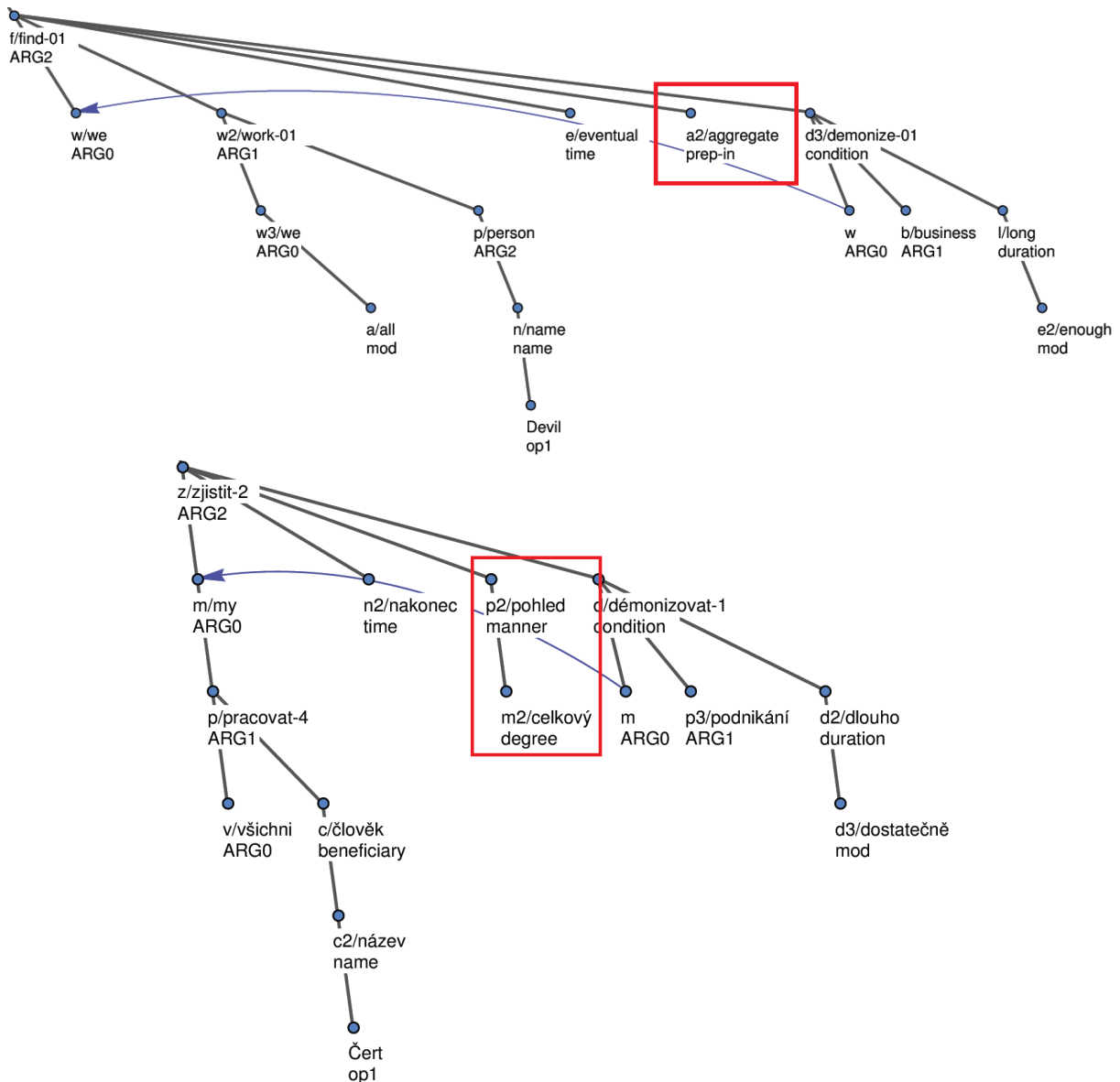
Figure 5: AMR structural difference: *in the aggregate* vs. *z/from celkového/overall pohledu/view to/it je/is tak/so, že/that ...*

deleting it was (probably) considered loss of information. It is only the confrontation with the translation to a different language when one realizes that with just a little more abstraction, the annotation could have been structurally the same (by keeping only the *commute* node in in the English annotation).[6]

Translation by interpretation is typically discouraged in translation school education, but sometimes it is necessary to use it for smooth understanding of the translated text. Often, such interpretation results in different AMR annotation. For example, *Virginia centrist* has been translated as *středový/centrist volič/voter [z/from Virginie/Virginia]*, because without the extra word *volič*, the literal translation of *centrist* would not be understandable correctly in this context (Fig. 6). Similarly, *a 55mph zone* vs. *zóna/zone s/with omezením/restriction na/to 55 mph* (added word *omezením/restriction*), or *traffic* vs. *dopravní/traffic zácpa/jam*.

Convention differences are inherent in many annotation schemes, and we have found them in AMR

---

[6]One could perhaps also argue that *adding* the equivalent of *back and forth* to the Czech translation would unify the translation, too; however, adding its literal equivalent in Czech *tam a zpět* would be considered superfluous and unnatural by Czech speakers.
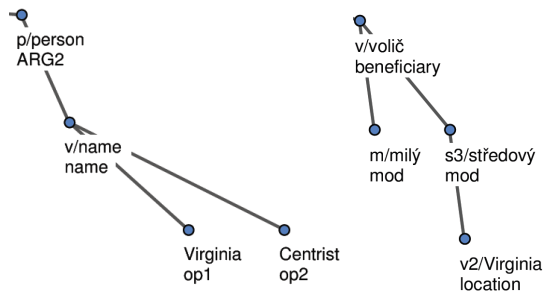
Figure 6: AMR structural difference: *Virginia Centrist* vs. *středový/centrist volič/voter [z/from Virginie/Virginia]*
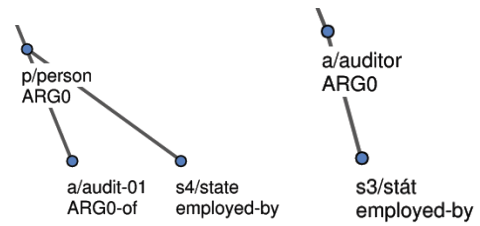
Figure 7: AMR convention difference: *auditor* as a single node vs. person, who audits

guidelines, too. Often, they were related to the use of `ARG-of` vs. keeping the nominalization as a single node. For example, for *auditor*, translated quite literally as *auditor* into Czech, has been annotated as "a person, who audits" in English while in the Czech AMR structure, there is a single node (Fig. 7) labeled as `auditor` (which undoubtedly will be correctly linked to some ontology entry after such linkage/wikification is complete). These differences might be harder to consolidate, since such conventions are very difficult to create proper guidelines for, especially across languages. No ontology (whether for events or objects) will be complete either (to base the decisions on a particular ontology content).

5   Conclusions and Future Work

We have investigated differences in the annotation of parallel texts using the Abstract Meaning Representation scheme, on approx. 1500 words of English-Czech corpus (100 sentences). We found and counted the number of identities and four types of differences (structural, structural local, relational, and referential), and exemplified them to see if a reconciliation (either by possibly changing the translation, the guidelines, or the annotation itself) is possible.

This is a work in progress. Substantial amount of work remains. We will have to use larger data, multiple annotation (interannotator agreement on English was relatively low and we expect to be the case on Czech, too, once two annotators start annotating the same sentences), and we would also have to actually suggest changes in the guidelines or their conventions, and to test them also on substantial amounts of data.

The immediate extension of this work will cover wikification, i.e. the linking of all nodes in the AMR representation of our dataset to some ontology: events are already covered, internally defined relations are already annotated, too (such as named entity types, dates, quantities, etc.), but external links remain to be added. We will not only use Wikipedia (as the term "wikification" might suggest), but we will extend this idea also to other sources, such as DBpedia or BabelNet, keeping all links in parallel if possible. This should allow for deep comparison of the two languages also content-wise. We should then be able to better answer the question of annotation unification which does depend on these links rather than on the annotation guidelines themselves.

Parallel AMR-annotated data will be used at the JHU 2014 Summer Workshop, where technology for AMR-based parsing, generation and possibly also MT will be developed, allowing also technological insight into the AMR scheme across languages.

Acknowledgements

---

[7] http://lindat.cz, resource used: http://hdl.handle.net/11858/00-097C-0000-0023-4338-F, also at http://lindat.mff.cuni.cz/services/PDT-Vallex

References

L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. 2013. Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop*, Sophia, Bulgaria.

Bernd Bohnet, Joakim Nivre, Igor Boguslavsky, Richard Farkas, Filip Ginter, and Jan Hajič. 2013. Joint morphological and syntactic analysis for richly inflected languages. *Transactions of the Association for Computational Linguistics*, 1:415–428.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 132–139. Association for Computational Linguistics.

Michael Collins. 1999. *Head-driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.

Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová, and Petr Pajas. 2003. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, volume 9 of *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*, pages 57–68. Växjö University Press, November 14–15, 2003.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In Jan Hajič, editor, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pages 1–18, Boulder, CO, USA. Association for Computational Linguistics.

Jan Hajič, Alena Böhmová, Eva Hajicová, and Barbora Hladká. 2003. The Prague Dependency Treebank: A Three Level Annotation Scenario. In Anne Abeillé, editor, *Treebanks: Building and Using Annotated Corpora*. Kluwer Academic Publishers.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.

A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The NomBank Project: An Interim Report. In *Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts.

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

Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *HLT-NAACL*, pages 404–411.

Sameer S Pradhan, Wayne Ward, Kadri Hacioglu, James H Martin, and Daniel Jurafsky. 2004. Shallow semantic parsing using support vector machines. In *HLT-NAACL*, pages 233–240.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.

James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, David Day, Lisa Ferro, Robert Gaizauskas, Marcia Lazo, Andrea Setzer, and Beth Sundheim. 2003. The TimeBank Corpus. *Corpus Linguistics*, pages 647–656.

Roser Saurí and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.

Jana Sindlerova, Zdenka Uresova, and Eva Fucikova. 2014. Resources in Conflict: A Bilingual Valency Lexicon vs. a Bilingual Treebank vs. a Linguistic Theory. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2490–2494, Reykjavik, Iceland, May 26-31. European Language Resources Association (ELRA).

Vivek Srikumar and Dan Roth. 2013. Modeling semantic relations expressed by prepositions. *Transactions of the Association for Computational Linguistics*, 1:231–242.

Zdeňka Urešová and Petr Pajas. 2009. Diatheses in the Czech Valency Lexicon PDT-Vallex. In Jana Levická and Radovan Garabík, editors, *Slovko 2009, NLP, Corpus Linguistics, Corpus Based Grammar Research*, pages 358–376, Bratislava, Slovakia. Jazykovedný ústav udovíta Štúra Slovenskej akadémie vied, Slovenská akadémia vied.

Zdeňka Urešová, 2006. *Verbal Valency in the Prague Dependency Treebank from the Annotator's Viewpoint*, pages 93–112. Veda, Bratislava, Bratislava, Slovensko.

Zdeňka Urešová. 2009. Building the PDT-VALLEX valency lexicon. In Catherine Smith Michaela Mahlberg, Victorina Gonzalez-Diaz, editor, *Proceedings of the Corpus Linguistics Conference)*, http://ucrel.lancs.ac.uk/publications/cl2009/100_FullPaper.doc, July 20-23. University of Liverpool, UK.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *EMNLP*, pages 88–94.

Nianwen Xue, Fei Xia, Fu dong Chiou, and Martha Palmer. 2005. The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, 11(2):207–238.

Nianwen Xue, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova, and Xiuhong Zhang. 2014. Not an Interlingua, But Close: Comparison of English AMRs to Chinese and Czech. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1765–1772, Reykjavik, Iceland, May 26-31. European Language Resources Association (ELRA).