Czech Dataset for Complex Aspect-Based Sentiment Analysis Tasks

Jakub Šmíd[1], **Pavel Přibáň**[1], Ondřej Pražák[1,2] and Pavel Král[1,2]

[1] Department of Computer Science and Engineering, Faculty of Applied Sciences,
[2] NTIS – New Technologies for the Information Society, Faculty of Applied Sciences,
University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic

E-mail: jaksmid@kiv.zcu.cz Web: nlp.kiv.zcu.cz

Motivation & Goals

- **Motivation**
  - Czech ABSA datasets do not allow evaluation of advanced ABSA tasks (TASD, ACTE and APD).
  - Existing datasets in a unified format in several other languages.
- **Goals**
  - Provide a reliable Czech ABSA dataset for compound ABSA tasks.
  - Align the annotation with the existing datasets in other languages.
  - Allow cross-lingual experiments and evaluation.

Dataset Construction

- Restaurant domain.
- APD, ACTE and TASD tasks.
- Two parts:
  - Manually annotated 3,189 reviews.
  - 24M additional raw reviews (330MB text).

Unsupervised Dataset

1. Obtain a list of names of Czech restaurants from Restu.cz.
2. Download restaurant reviews from Google Maps during September 2022.

Dataset Annotation

- Manual annotation by 3 Czech native speakers.
- **Two data segments**:
  1. Reused reviews from existing Czech dataset (Hercig et al., 2016) in SemEval-2014 format.
  2. Randomly sampled 1,110 reviews from the unsupervised dataset.
- Reviews reannotated into SemEval-2016 format.

1. **Identify objective reviews:** Objective reviews and reviews without any sentiment expressed marked as *"OutOfScope"*.
2. **Identify aspect terms:** One or more word expressions naming a specific aspect of the target entity, e.g. *"toast with eggs"*.
3. **Assign aspect category:** Assign aspect categories for each identified aspect term. The aspect category consists of entity and attribute ($E\#A$) and must be chosen from a pre-defined set of categories (e.g. $RESTAURANT\#GENERAL$ or $FOOD\#QUALITY$).
4. **Assign sentiment polarity:** For each (aspect term, aspect category) tuple assign the sentiment polarity from one of the following values: *neutral*, *negative*, and *positive*.

Figure 1: Our annotations compared with Hercig et al. (2016).

Inter-annotator Agreement

Annotation target	IAA
Aspect term	93.19
Aspect category	93.00
Aspect term & aspect category	91.06
Aspect term & aspect category & polarity	89.70

Table 2: Inter-annotator agreement (IAA) as micro F1 score.

References

Hercig, T., Brychcín, T., Svoboda, L., Konkol, M., & Steinberger, J. (2016). Unsupervised methods to improve aspect-based sentiment analysis in czech. *Computación y Sistemas*, *20*(3), 365–375.

Aspect-Based Sentiment Analysis - Tasks

- Fine-grained sentiment analysis.
- Multiple sentiment elements.

"*Delicious steak*"

⇒ aspect term (a) – "*steak*"

⇒ aspect category (c) – "*food quality*"

⇒ sentiment polarity (p) – "*positive*"

Task	Input	Output	Example output
ATE	s	$\{a\}$	{"steak", "water"}
ACD	s	$\{c\}$	{food, drinks}
APD	s, "steak", food	p	POS
E2E-ABSA	s	$\{(a, p)\}$	{("steak", POS), ("water", NEG)}
ACTE	s	$\{(a, c)\}$	{("steak", food), ("water", drinks)}
TASD	s	$\{(a, c, p)\}$	{("steak", food, POS), ("water", drinks, NEG)}

Table 1: Input and output format for ABSA tasks for a review s: *"Delicious steak but expensive water"*.

Dataset Details

- Three versions of our dataset – `CsRest`.
  - `CsRest-O` exclusively comprises the reannotated data from Hercig et al. (2016).
  - `CsRest-N` – all annotated data, all the new data (not present in the original data from Hercig et al., 2016) serves as the test data.
  - `CsRest-M` – all annotated data, 25% randomly selected as test data.
- Second largest ABSA dataset.

Split	Count	CsRest-O	CsRest-N	CsRest-M
Train	Reviews	1,450	1,934	2,151
	Triplets	2,510	3,240	4,386
	No triplets	104	142	109
	NULL terms	590	795	961
Dev	Reviews	162	215	240
	Triplets	253	430	483
	No triplets	6	17	9
	NULL terms	64	95	104
Test	Reviews	537	1,040	798
	Triplets	907	2,808	1,609
	No triplets	49	0	41
	NULL terms	49	517	342
Total	Reviews	2,149	3,189	3,189
	Triplets	3,670	6,478	6,478
	No triplets	159	159	159
	NULL terms	890	1,407	1,407

Table 3: Statistics of our dataset.

Dataset & Source Code

https://github.com/biba10/Czech-Dataset-for-Complex-ABSA-Tasks

Cross-language Comparison

Dataset	Reviews	Triplets
en	2,676	3,366
es	2,951	3,792
fr	2,429	5,322
nl	2,286	2,473
ru	4,699	5,322
tr	1,248	1,694
CsRest-O	2,149	3,670
CsRest-N	3,189	6,478
CsRest-M	3,189	6,478

Table 4: Comparison of restaurant datasets across languages.

Sequence-to-Sequence Models

- Input and output text.
- Multiple ABSA tasks simultaneously.
- Annotations converted to natural language:

c is $P_p(p)$, given the expression: a

$$P_p(p) = \begin{cases} great & \text{if } p \text{ is } positive, \\ ok & \text{if } p \text{ is } neutral, \\ bad & \text{if } p \text{ is } negative. \end{cases}$$

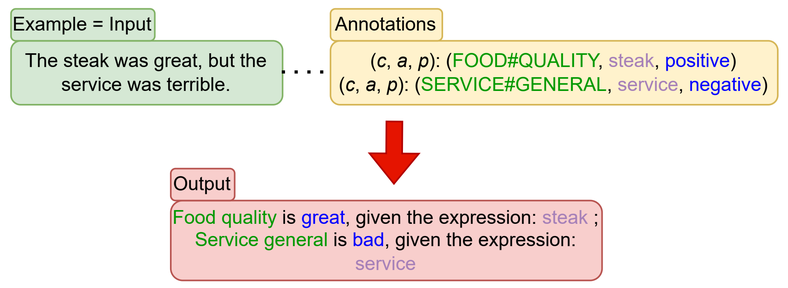

Figure 2: Sequence-to-sequence approach.

Model	Task	CsRest-O	CsRest-N	CsRest-M
mT5	ACD	$75.4^{\pm 1.8}$	$68.9^{\pm 1.1}$	$70.8^{\pm 1.5}$
	ATE	$66.5^{\pm 2.5}$	$59.7^{\pm 1.5}$	$66.9^{\pm 1.4}$
	ACTE	$56.4^{\pm 1.0}$	$45.0^{\pm 1.7}$	$52.6^{\pm 1.8}$
	TASD	$48.0^{\pm 1.0}$	$41.1^{\pm 1.8}$	$46.4^{\pm 1.5}$
mBART	ACD	$\mathbf{78.7}^{\pm 1.6}$	$\mathbf{79.3}^{\pm 0.4}$	$\mathbf{80.6}^{\pm 1.7}$
	ATE	$\mathbf{78.9}^{\pm 1.3}$	$\mathbf{76.0}^{\pm 1.5}$	$\mathbf{79.7}^{\pm 1.1}$
	ACTE	$\mathbf{67.2}^{\pm 1.4}$	$\mathbf{62.4}^{\pm 0.7}$	$\mathbf{67.3}^{\pm 1.2}$
	TASD	$\mathbf{57.5}^{\pm 1.7}$	$\mathbf{56.3}^{\pm 0.6}$	$\mathbf{59.3}^{\pm 1.4}$

Table 6: Results for sequence-to-sequence models.

Encoder-Based Models

- Czech and multilingual models.
- Additional pre-training on unsupervised data.

Model	CsRest-O				CsRest-N				CsRest-M			
	APD	ACD	ATE	E2E	APD	ACD	ATE	E2E	APD	ACD	ATE	E2E
Czert	$83.2^{\pm 1.4}$	$81.2^{\pm 1.4}$	$81.7^{\pm 0.4}$	$66.8^{\pm 0.7}$	$85.5^{\pm 4.9}$	$81.6^{\pm 1.5}$	$78.4^{\pm 1.0}$	$70.9^{\pm 1.2}$	$85.3^{\pm 0.9}$	$82.2^{\pm 0.3}$	$82.8^{\pm 0.7}$	$70.6^{\pm 0.9}$
RobeCzech	$85.2^{\pm 1.6}$	$80.9^{\pm 2.5}$	$82.9^{\pm 0.4}$	$67.8^{\pm 1.6}$	$89.4^{\pm 1.2}$	$80.8^{\pm 1.6}$	$78.8^{\pm 1.1}$	$71.9^{\pm 1.6}$	$87.6^{\pm 1.3}$	$83.1^{\pm 1.0}$	$82.8^{\pm 0.5}$	$71.3^{\pm 1.9}$
FERNET	$\underline{86.0}^{\pm 0.4}$	$\underline{83.7}^{\pm 1.2}$	$\mathbf{84.9}^{\pm 1.1}$	$\underline{71.7}^{\pm 2.1}$	$\underline{90.1}^{\pm 2.2}$	$\mathbf{82.9}^{\pm 0.9}$	$\mathbf{80.8}^{\pm 1.2}$	$\underline{74.7}^{\pm 1.5}$	$\mathbf{88.2}^{\pm 0.8}$	$\underline{84.3}^{\pm 0.4}$	$\underline{83.2}^{\pm 1.1}$	$\mathbf{74.8}^{\pm 1.1}$
Small-E-Czech	$78.0^{\pm 3.7}$	$75.5^{\pm 1.2}$	$81.5^{\pm 0.9}$	$59.2^{\pm 4.8}$	$84.6^{\pm 1.6}$	$76.7^{\pm 2.0}$	$77.3^{\pm 2.4}$	$64.0^{\pm 2.9}$	$83.3^{\pm 0.8}$	$79.8^{\pm 0.7}$	$81.1^{\pm 0.7}$	$66.7^{\pm 2.0}$
mBERT	$77.1^{\pm 3.8}$	$77.8^{\pm 2.1}$	$79.6^{\pm 0.6}$	$60.3^{\pm 2.8}$	$85.1^{\pm 1.8}$	$78.6^{\pm 1.3}$	$76.2^{\pm 1.7}$	$67.5^{\pm 2.0}$	$82.2^{\pm 1.0}$	$79.0^{\pm 0.5}$	$80.0^{\pm 0.6}$	$67.7^{\pm 1.6}$
XLM-R_{BASE}	$80.7^{\pm 2.3}$	$80.4^{\pm 1.4}$	$82.4^{\pm 0.6}$	$68.9^{\pm 3.6}$	$88.5^{\pm 1.8}$	$80.6^{\pm 1.7}$	$78.6^{\pm 1.3}$	$70.7^{\pm 2.1}$	$85.1^{\pm 1.6}$	$81.0^{\pm 1.1}$	$82.0^{\pm 0.4}$	$70.4^{\pm 0.7}$
XLM-R_{LARGE}	$\mathbf{87.2}^{\pm 1.5}$	$\mathbf{85.7}^{\pm 0.4}$	$\underline{84.0}^{\pm 0.8}$	$\mathbf{71.9}^{\pm 2.3}$	$\mathbf{91.4}^{\pm 0.9}$	$\underline{82.8}^{\pm 1.0}$	$\underline{80.2}^{\pm 1.1}$	$\mathbf{75.5}^{\pm 1.0}$	$\underline{87.9}^{\pm 0.8}$	$\mathbf{86.2}^{\pm 0.3}$	$\mathbf{83.5}^{\pm 1.2}$	$\underline{74.4}^{\pm 1.0}$
Czert*	$88.4^{\pm 0.7}$	$86.8^{\pm 0.9}$	$85.7^{\pm 1.7}$	$74.7^{\pm 1.4}$	$89.2^{\pm 2.6}$	$84.6^{\pm 0.4}$	$81.3^{\pm 1.4}$	$73.8^{\pm 1.2}$	$88.3^{\pm 1.1}$	$86.1^{\pm 0.5}$	$84.4^{\pm 1.0}$	$75.6^{\pm 0.5}$
RobeCzech*	$88.4^{\pm 0.9}$	$84.9^{\pm 0.7}$	$85.3^{\pm 1.1}$	$70.4^{\pm 2.3}$	$91.1^{\pm 0.8}$	$83.9^{\pm 0.7}$	$82.3^{\pm 1.0}$	$74.3^{\pm 0.7}$	$88.4^{\pm 0.8}$	$85.7^{\pm 0.9}$	$84.9^{\pm 1.2}$	$75.4^{\pm 1.1}$
FERNET*	$85.0^{\pm 1.1}$	$83.9^{\pm 0.7}$	$84.0^{\pm 0.9}$	$71.7^{\pm 1.0}$	$91.0^{\pm 1.5}$	$84.0^{\pm 1.5}$	$82.3^{\pm 1.2}$	$75.9^{\pm 0.8}$	$90.0^{\pm 0.5}$	$87.1^{\pm 0.4}$	$85.6^{\pm 0.8}$	$77.0^{\pm 0.4}$

Table 5: Results with encoder-based models. * denotes models with additional pre-training.

Error Analysis

- Occasional problems with output format for mT5.
- Aspect term prediction – part completion (e.g. *"burrito"* instead of *"burrito bowl"*), correcting typos (e.g. *"sevrice"* to *"service"*), making up words not presented in the text.
- Confusion of similar categories.
- Neutral sentiment polarity.

Conclusion

- Novel manually annotated Czech ABSA dataset in the restaurant domain.
- Possible to solve complex ABSA tasks by linking sentiment elements together.
- Unlabelled corpora for unsupervised training.
- Strong baseline results for various ABSA tasks using Transformer-based models.