

Turkish Universal Conceptual Cognitive Annotation

Necva Bölücü¹, Burcu Can²

¹ Hacettepe University, Department of Computer Engineering

² University of Wolverhampton, Research Group in Computational Linguistics

LREC 2022

Table of Contents

- 1 Introduction
- 2 Datasets
- 3 Universal Conceptual Cognitive Annotation (UCCA)
- 4 Turkish UCCA Dataset
- 5 Experiments & Results
- 6 Conclusions

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Introduction

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Introduction

- Semantic representation is a way of mapping a given text into its formal or abstract form.
- Graph-based semantic representations:
 - Abstract Meaning Representation (AMR) [1]
 - Universal Conceptual Cognitive Annotation (UCCA) [2]
 - Bilexical Semantic Dependencies (SDP) [3]
 - Universal Decompositional Semantics (UDS) [4]

Annotation

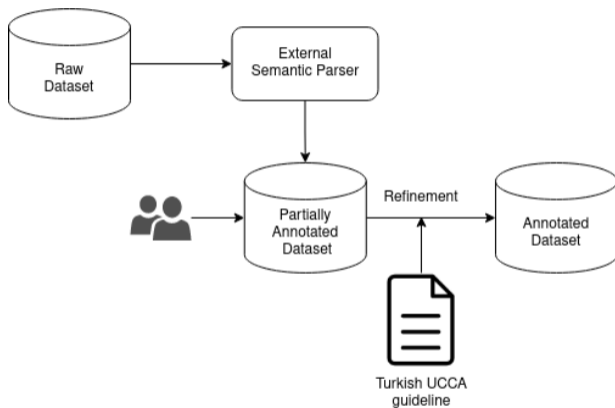


Figure: Turkish UCCA dataset annotation procedure

Datasets

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Datasets

- Semantically Annotated Datasets in Turkish
 - Turkish Abstract Meaning Representation (AMR) [5]
- UCCA-annotated Datasets in Other Languages
 - English UCCA dataset [2]
 - German UCCA dataset [6]
 - French UCCA dataset [7]

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Universal Conceptual Cognitive Annotation (UCCA)

Turkish UCCA

Necva Bölücü

Introduction

Datasets

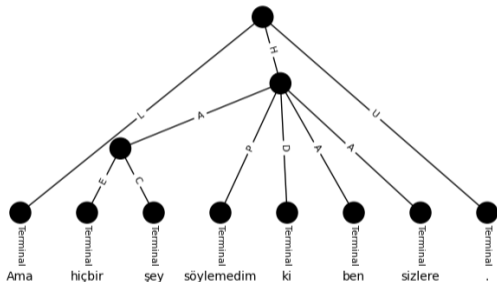
Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

What is Universal Conceptual Cognitive Annotation?



Categories:

Scene elements:

- Process (P)
- State (S)
- Participant (A)
- Adverbial(D)

Non-Scene elements:

- Center (C)
- Elaborator (E)
- Connector (N)
- Relator (R)

Inter-Scene relations:

- Parallel Scene (H)
- Linker (L)
- Ground (G)

Other:

- Function (F)

Figure: The UCCA annotation graph of the Turkish sentence “Ama hiçbir şey söylemedim ki sizlere.” (in English, “But I did not say anything to you.”)

Turkish UCCA Dataset

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Turkish Annotation Rules

- **Pronoun-dropping**

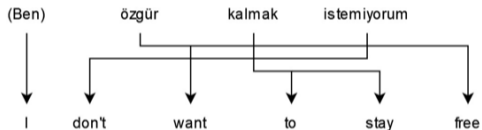


Figure: $(Ben)_{A-IMPLICIT}$ özgür kalmak istemiyorum (in English, “I don't want to stay free”)

Turkish Annotation Rules

- **Genitive case**

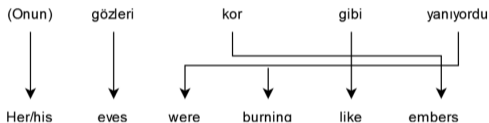


Figure: $(Onun)_{E-IMPLICIT}$ *gözleri kor gibi yanıyordu* (in English, “Her/his eyes were burning like embers”)

Turkish Annotation Rules

- Postpositions

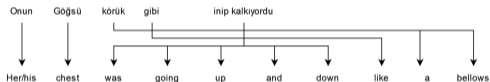


Figure: *Göğsü körük gibi_R inip kalkıyordu* (in English, “*Her/his chest was going up and down like a bellows*”)

Turkish Annotation Rules

- **Negation**

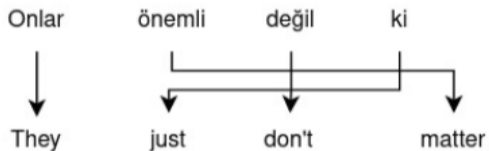


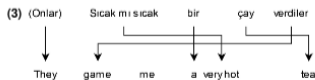
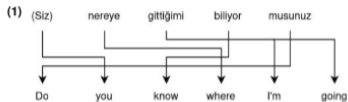
Figure: *Onlar önemli değil_D ki* (in English, “*They just don't matter*”)

Turkish Annotation Rules

- **Clitics:**

- “mi”

- ① Yes/No condition:
- ② Adverbial clause marker:
- ③ Intensifier in doubled forms:

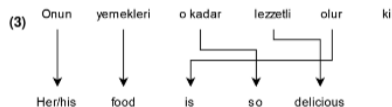
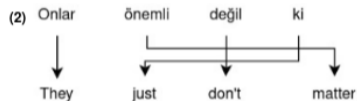
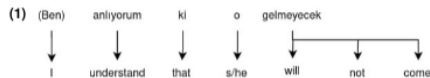


Turkish Annotation Rules

- **Clitics:**

- **“ki”**

- 1 Subordinator connecting:
- 2 Repudiative discourse connective
- 3 Exclamations:
- 4 Relative clause marker:

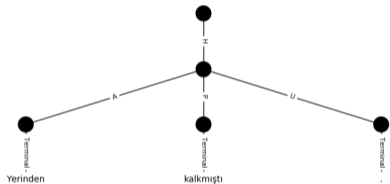


Inter-Annotator Agreement

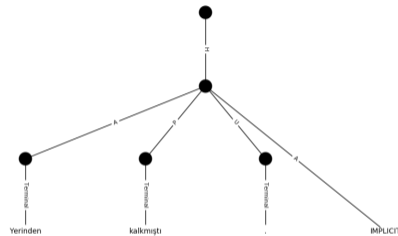
- 2 annotators who are native speakers of Turkish refined the partially annotated dataset.
- Agreement between annotators is 90.69%¹
- Inter-annotator Agreement (IAA) (Cohen's Kappa [8]) is 89.05

¹There was disagreement in 38 out of 408 tokens in 50 sentences between the two annotators during annotation.

Comparison of Annotations - 1



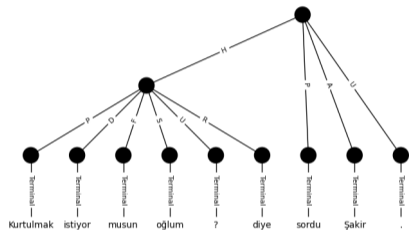
(a) The output of the semantic parsing model



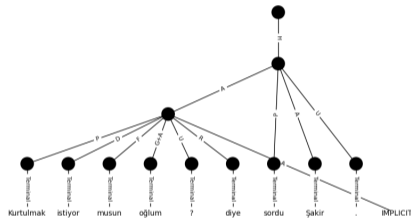
(b) The correct annotation of the sentence

Figure: The semantic parse tree obtained from the semantic parsing model and the gold standard annotation obtained from the manual annotation of the sentence, *Yerinden kalkmıştı.* (in English, "S/he had gotten up.")

Comparison of Annotations - 2



(a) The output of the semantic parsing model



(b) The correct annotation of the sentence

Figure: The semantic parse tree obtained from the semantic parsing model and the gold standard annotation obtained from the manual annotation of the sentence, *Kurtulmak istiyor musun oğlum? diye sordu Şakir.* (in English, "Do you want to be saved, son? asked Shakir.")

Annotation Statistics

	Turkish	En-Wiki	En-20K
# edges	407	208,937	16,803
% primary	96.57	97.40	96.79
% remote	3.43	2.60	3.21
% Participant (A)	28.19	17.17	18.1
% Center (C)	8.09	18.74	16.31
% Adverbial (D)	8.82	3.65	5.25
% Elaborator (E)	4.41	18.98	18.06
% Function (F)	3.19	3.38	3.58
% Ground (G)	0.98	0.03	0.56
% Parallel Scene (H)	7.11	6.02	6.3
% Linker (L)	1.23	2.19	2.66
% Connector (N)	0.00	1.26	0.93
% Process (P)	15.44	7.1	7.51
% Relator (R)	2.70	8.58	8.09
% State (S)	4.17	1.62	2.1
% Punctuation (U)	15.69	11.28	10.55

Table: Statistics of the UCCA datasets in Turkish and English. Proportions of the edges and labels of each dataset are given.

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Experiments & Results

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Dataset

	English-Wiki	English-20K	German-20K	French-20K
Train	4,113	0	5,211	15
Validation	514	0	651	238
Test	515	492	652	239

Table: The number of sentences in each UCCA-annotated dataset provided by SemEval 2019 [6]

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Results

		Primary			Remote			Average		
		P	R	F_1	P	R	F_1	P	R	F_1
Zero-shot	Labeled	73.55	80.25	76.76	0.00	0.00	0.00	73.04	77.89	75.39
	Unlabeled	76.80	83.80	80.15	0.00	0.00	0.00	76.80	81.33	79.00
Few-shot	Labeled	78.72	84.30	81.42	30.00	25.00	27.28	78.75	83.78	81.11
	Unlabeled	87.71	93.92	90.71	50.00	41.67	45.45	87.53	93.12	90.24

Table: Precision (P), Recall (R) and F-measure (F_1) results obtained from zero-shot and few-shot learning on the Turkish UCCA dataset

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Conclusions

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Conclusions

- We started annotation first Turkish UCCA dataset with 50 sentences
- We defined semi-automatic pipeline for annotation
- We defined new rules for discrepancies between the annotated sentences and the English guideline.
- We performed experiments with zero-shot and few-shot learning to evaluate the new annotated dataset.

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

Future Work

- We aim to extend the dataset with the same approach.
- We will also create a more comprehensive guideline for Turkish annotation.

Turkish UCCA

Necva Bölücü

Introduction

Datasets




Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

References I

-  L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider, “Abstract meaning representation for sembanking,” in *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pp. 178–186, 2013.
-  O. Abend and A. Rappoport, “UCCA: A semantics-based grammatical annotation scheme,” in *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers*, pp. 1–12, 2013.
-  S. Oepen, M. Kuhlmann, Y. Miyao, D. Zeman, S. Cinková, D. Flickinger, J. Hajic, A. Ivanova, and Z. Uresova, “Towards comparability of linguistic graph banks for semantic parsing,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pp. 3991–3995, 2016.

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

References II

-  A. S. White, D. Reisinger, K. Sakaguchi, T. Vieira, S. Zhang, R. Rudinger, K. Rawlins, and B. Van Durme, “Universal decompositional semantics on universal dependencies,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1713–1723, 2016.
-  Z. Azin and G. Eryiğit, “Towards Turkish Abstract Meaning Representation,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 43–47, 2019.
-  D. Hershcovich, Z. Aizenbud, L. Choshen, E. Sulem, A. Rappoport, and O. Abend, “SemEval-2019 Task 1: Cross-lingual Semantic Parsing with UCCA,” in *Proceedings of the 13th International Workshop on Semantic Evaluation*, pp. 1–10, 2019.

Turkish UCCA

Necva Bölücü

Introduction

Datasets



Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions

References III

-  E. Sulem, O. Abend, and A. Rappoport, “Conceptual Annotations Preserve Structure Across Translations: A French-English Case Study,” in *Proceedings of the 1st Workshop on Semantics-Driven Statistical Machine Translation (S2MT 2015)*, pp. 11–22, 2015.
-  J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

Turkish UCCA

Necva Bölücü

Introduction

Datasets

Universal
Conceptual
Cognitive
Annotation
(UCCA)

Turkish UCCA
Dataset

Experiments
& Results

Conclusions