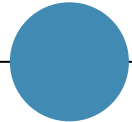


# Defining a Framework for Semantic Categories for Turkish Nominal Morphemes

Yağmur Öztürk<sup>1</sup>, Izabella Thomas<sup>1</sup>, Snejana Gadjeva<sup>2</sup>

<sup>1</sup>CRIT, Université de Franche-Comté


<sup>2</sup>CREE, Institut National des Langues et Civilisations Orientales



# CONTENT

 Introduction

 Computerised and linguistic studies

 Methodology - Semantic categories

 Our computerised resource Semantürk

 Conclusion

# Turkish derivational morphology

→ Agglutinative – suffixation

(1) Turkish	English
göz	eye
göz-lük	eyeglasses
göz-lük-çü	optician
göz-lük-çü-lük	opticianry

# Semantic description of morphemes in Turkish

- Morphology and semantics with nominals under study in many languages (Bagasheva, 2017; Namer et al., 2019; etc)

What about Turkish nominals ?

"[A] general area with no or little resources is **semantics.**" Çöltekin et. al (2023)

---

# Morphological analysers

Available resources:

- Zemberek (Akin and Akin, 2007) – Java Library
- TRmorph (Çöltekin, 2014) – Foma FST
- Turkish Morphological Analyzer (Yıldız et al., 2019) – Java Library
- Trnlp (Bayol, 2018) – Python Library

# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

göz +

*en. eye*

# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

göz +lük

*en. eyeglasses*



# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

göz +lük<sub>ARTIFACT</sub>

*en. eyeglasses*

# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

göz +lük<sub>ARTIFACT</sub> +çü<sub>AGENT</sub>

*en. optician*

# Morphological analysers

Extract 1 from TRmorph (Çöltekin, 2014):

Analysis for the word **gözlükçülük**:

- gözlükçü<N><lik><Adj>
- gözlükçü<N><lik><N>
- gözlük<N><ci><N><lik><N>
- gözlük<N><ci><N><lik><Adj>
- göz<N><lik><N><ci><N><lik><N>

göz +lük<sub>ARTIFACT</sub> +çü<sub>AGENT</sub> +lük<sub>PROFESSION</sub>

*en. opticianry*

---

# Linguistics resources

Semantic concepts in affixation (Bagasheva, 2017)

- ✓ **Local** and **comparative** research
- ✓ Applied to derivational **affixes**
- ✓ All types of derivation, including **nominal**
- ✓ Different levels of **granularity**
- ✓ Cross-linguistic project on 40 languages (Körtvélyessy et al., 2020) including **Turkish** (Göksel & Kaplan, 2020)

Table 1. Set of comparative semantic concepts in affixation. The table illustrates English and, additionally, a different language (typically Bulgarian), if the concept does not apply in English and/or for illustration in a language other than English. The language from which the example has been extracted is in front of the word.

<b>Comparative semantic concept</b>	<b>Emergent meaning</b>	<b>Examples</b>		
ABILITY	Possibility to be processed in a particular way	Eng.	<i>readable</i>	<i>readability</i>
		Bul.	<i>četiven</i>	<i>četivnost</i>
			‘readable’	‘readability’
ABSTRACTION	Name of an abstract idea	Eng.	<i>justice</i>	
		Bul.	<i>pravda</i>	
			‘justice’	
ACTION	Performing of an activity	Eng.	<i>reading</i>	
		Bul.	<i>strelba</i>	
			‘shooting’	

# Linguistics resources

Tested on 50 derived Turkish nouns extracted from a Turkish online dictionary (*Türk Dil Kurumu sözlükleri*<sup>1</sup>):

<b><i>Types of semantic matches</i></b>	<b><i>Number of matches</i></b>
Matched with one	25
Matched with several	7
Unmatched	18

<sup>1</sup> <https://sozluk.gov.tr/>

# Linguistics resources

- Ambiguous semantic concept names and definitions

Comitative	Co-participant	Eng. <i>co-worker</i> Bul. <i>sâdružnik</i> 'partner'
------------	----------------	--

→ Inflectional marker (Göksel & Kerslake, 2005)

(2) kedi-**yle**

cat-COM

'with the cat'

- Non-explicit granularity
- Incompatible concepts for some morphemes' semantics

# Linguistics resources

- Ambiguous semantic concept names and definitions
- Non-explicit granularity

Dweller	Occupant of a specified location	Eng. <i>co-worker</i> Bul. <i>sâdružnik</i> 'partner'
Entity	Objectification; object that has real existence, material expression	Bul. <i>kostilka</i> 'pit, stone'

- Incompatible concepts for some morphemes' semantics



---

# Linguistics resources

- Ambiguous semantic concept names and definitions
- Non-explicit granularity
- Incompatible concepts for some morphemes' semantics

(3) meslek-taş  
profession-?  
'colleague'

---

# Linguistics resources

## WordNet

- ✓ can be adapted and used for morpho-semantic description
  - ❑ Semantic annotation by Huguin et al. (2022 – to be published) for the Demonext project (Namer. et al, 2019)
- ✓ developed for a variety of languages including Turkish (Bilgin et al., 2004; Çetinoğlu et al., 2018)

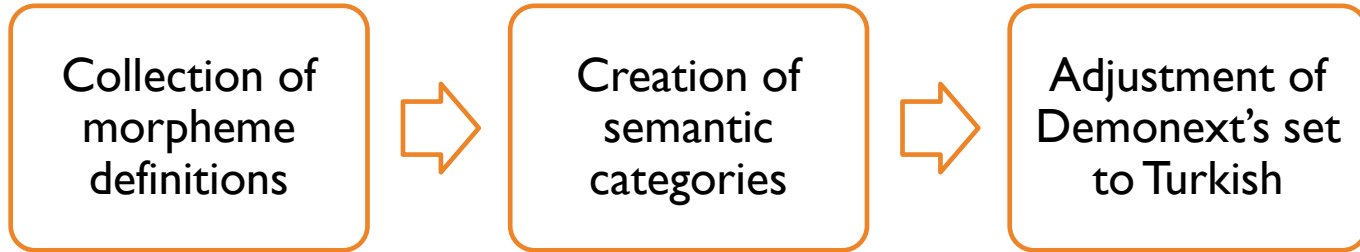
---

# Linguistics resources

- Semantic concept with extended definitions and linguistics tests to guide
- Explicit hierarchical structure with an ontological structure

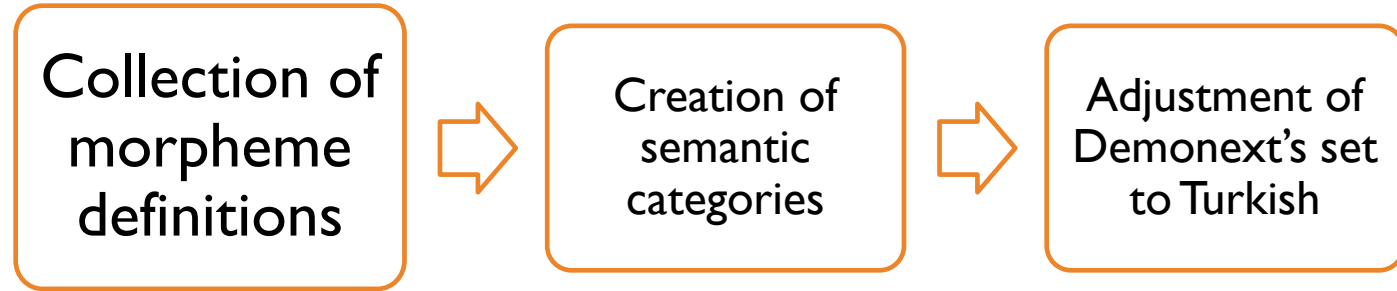
Adaptable to Turkish nominal morphosemantics ?

# Semantic categories



- check if the ontology captures all of the morphemes semantics

# Semantic categories



- ❑ Introductory study in French (Bazin, 1994),
- ❑ Turkish grammar books in Turkish (Adalı, 2004; Korkmaz, 2014; Boz, 2015) and in English (Göksel & Kerlake, 2005),
- ❑ article (Akçataş & Taşdemir, 2020) and
- ❑ textbooks (Bozdemir, 1991; Erikan et. al, 2008).

---

# Semantic categories

Difficulties:

- Different categorisation of morphemes

## 1. Nominals

- Nouns
- Adjectives
- Adverbs
- Pronouns

## 2. Verbs

- Inconsistencies between references
- Lack of formalised description

---

# Semantic categories

1. Selection of N-to-N morphemes from sources,
2. Stored any description given in a single file,
  - References
  - Morpheme
  - Definition
2. Aligned each definition found from each source by morpheme,
3. Checked if there is a match with an existing category from the top WordNet ontology

# Semantic categories

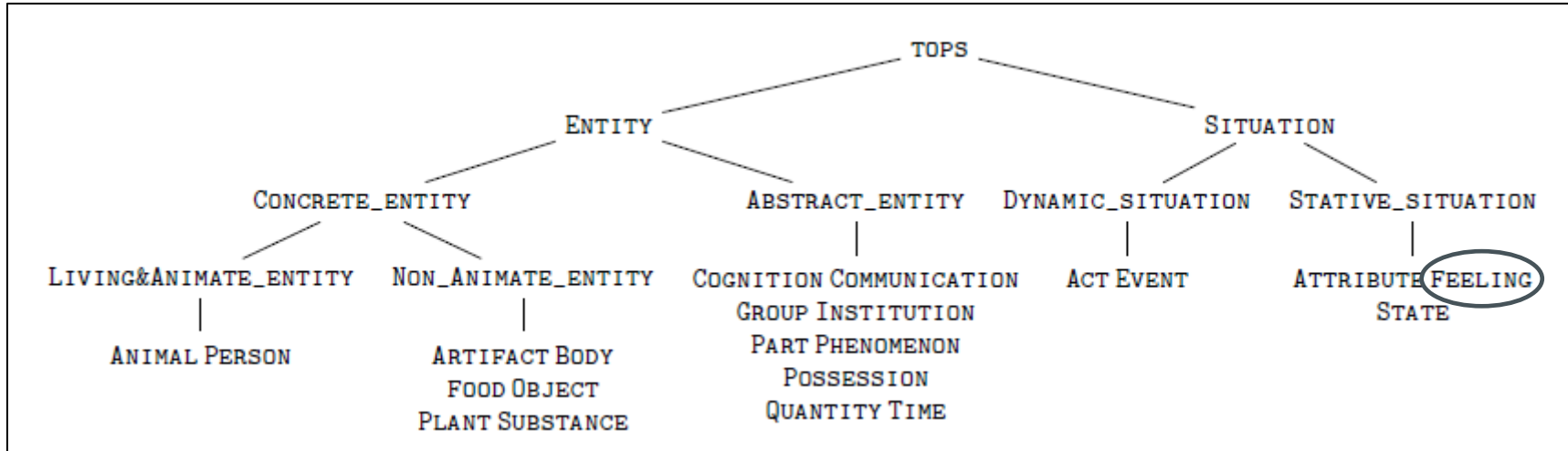
<b>Adalı (2004)</b>	-
<b>Korkmaz (2014)</b>	41. +Cağlz: +Cağlz eki, +CAk ekinin yine bir perkiştirme eki olan +Az > +Iz ekiyle genişletilmesinden oluşmuş bir birleşik ektir. Adlardan ve ad niteliğinde olan sözlerden küçültme, sevgi ve acıma ifadesi veren küçültme adları türetir: <i>adamcağız</i> , [...]
<b>Boz (2015)</b>	Küçültme, şefkat, acıma sevgi ve alay - iii. +{cAĞlz} işlekliliğini yitirmiş bir biçimbirimdir: <i>kız+cağız</i> , <i>kadın+cağız</i> , <i>tavuk+cağız</i>
<b>Akçataş ve Taşdemir (2020)</b>	+cAğız: Eklendiği kelimeye “acıma” hissi verir; <i>hayvancağız</i> , <i>yavrucağız</i> , <i>çocukcağız</i> ...



# Semantic categories

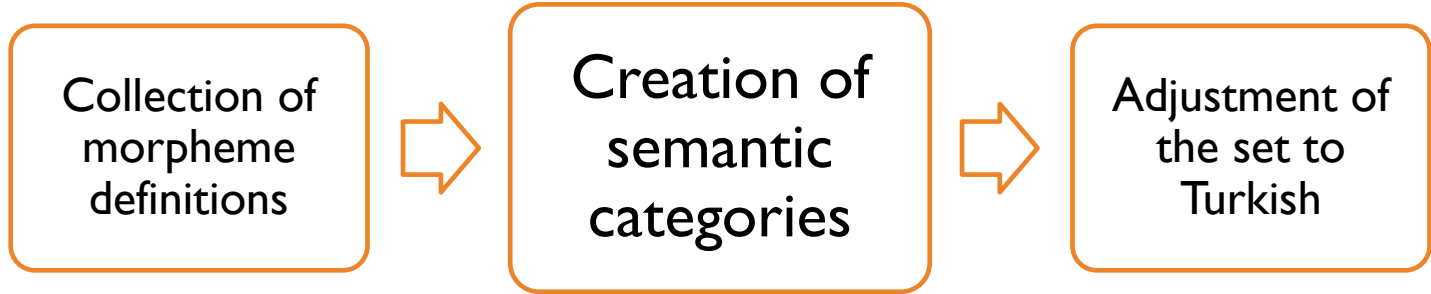
<b>Adalı (2004)</b>	-
<b>Korkmaz (2014)</b>	41. +Cağlz: +Cağlz eki, +CAk ekinin yine bir perkiştirme eki olan +Az > +Iz ekiyle genişletilmesinden oluşmuş bir birleşik ektir. Adlardan ve ad niteliğinde olan sözlerden <b>küçültme, sevgi ve acıma ifadesi veren küçültme adları türetir: adamcağız, [...]</b>
<b>Boz (2015)</b>	<b>Küçültme, şefkat, acıma sevgi ve alay</b> - iii. +{cAĞlz} işlekliliğini yitirmiş bir biçimbirimdir: <i>kız+cağız, kadın+cağız, tavuk+cağız</i>
<b>Akçataş ve Taşdemir (2020)</b>	+cAğız: Eklendiği kelimeye “ <b>acıma</b> ” hissi verir; <i>hayvancağız, yavrucağız, çocukcağız...</i>

# Which category matches the meaning added by -cAğlz ?



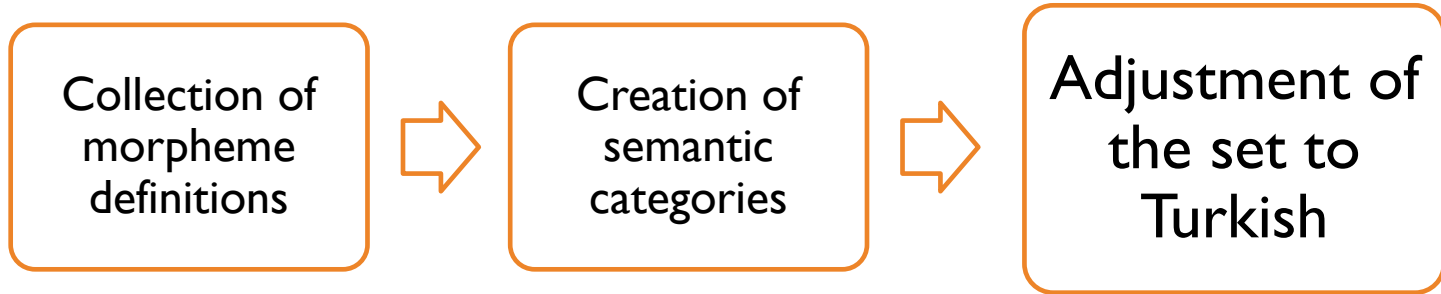
*22 simple categories organised in an ontological hierarchy designed specifically for nominal description (Huguin et. al, 2022 – to be published)*

# Semantic categories



Specific feeling → PITY

# Semantic categories

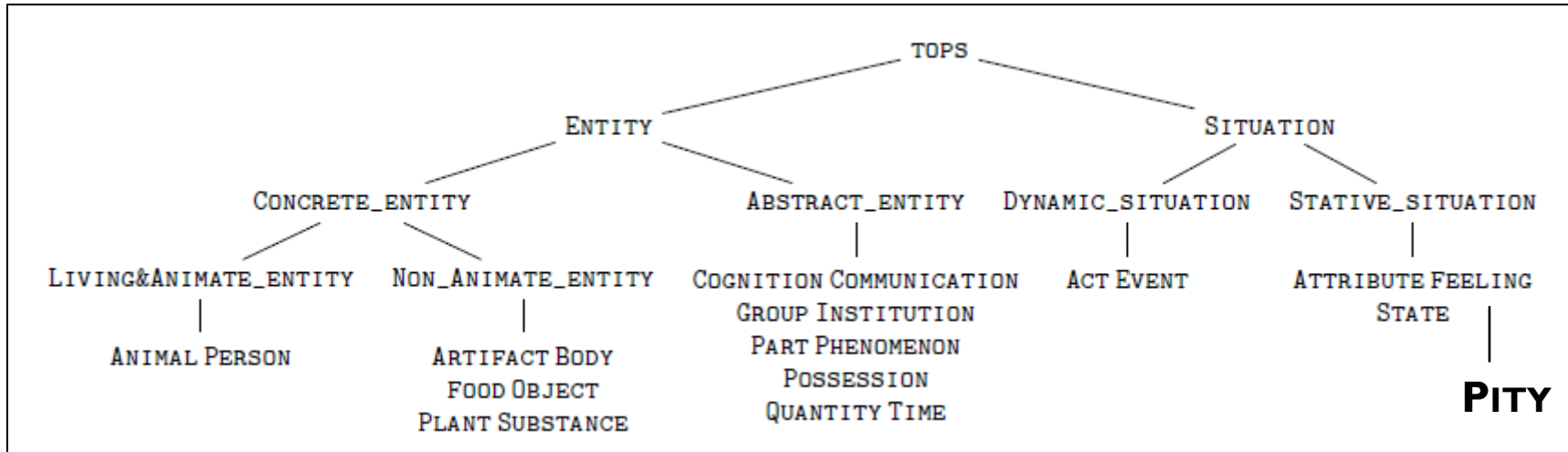


Specific feeling → PITY

FEELING

- temporary psychological or physiological state
- the subject is the experiencer of the denoted affect

# Which category matches the meaning added by -cAğlz ?



*22 simple categories organised in an ontological hierarchy designed specifically for nominal description (Huguin et. al, 2022 – to be published)*

---

# Semantürk's characteristics

-20 simple categories added to the structure

OWL / XML

Open Science Framework

- Available for any usage in linguistics and NLP
  - Coherent in its structure
  - Shared for different purposes
  - Understandable and readable
- Interoperable

---

# Semantürk's characteristics

- Usage in other languages
  - WordNet
  
- Extensible
  - Fixed structure
  - New categories

---

## To Do

- Adaptation of linguistic tests for each of the semantic categories, comprising the top categories
- Evaluation of the semantic categories by Turkish users
- Test the compatibility with a morphological analyser





## REFERENCES (SELECTED)

- Adalı, Oya. 2004. *Türkiye Türkçesinde Biçimbirimler*. Istanbul: Papatya.
- Akçataş, Ahmet, and Serpil Taşdemir. 2020. 'Türkiye Türkçesinde Kök Ya Da Gövdeye Gelen Ekler Üzerine Bir Anlambilim İncelemesi (a Semantic Study about Affixes on the Root and Stem of Turkish)'. *Avrasya Dil Eğitimi ve Araştırmaları Dergisi* 4 (1): 129–49.
- Bagasheva, A. 2017. « Comparative semantic concepts in affixation ». In: Salvador Valera & Juan Santana (eds.), *Competing Patterns in English Affixation*, 33–65. Peter Lang,
- Bilgin, Orhan, Özlem Çetinoğlu, and Kemal Oflazer. 2004. 'Building a Wordnet for Turkish'. *Romanian Journal of Information Science and Technology* 7 (1–2): 163–72.
- Çöltekin, Çağrı. 2014. 'A Set of Open Source Tools for Turkish Natural Language Processing'. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1079–86. Reykjavik, Iceland: European Language Resources Association (ELRA).
- Çöltekin, Çağrı, A. Seza Doğruöz, and Özlem Çetinoğlu. 2023. 'Resources for Turkish Natural Language Processing: A Critical Survey'. *Language Resources and Evaluation* 57 (1): 449–88.



## REFERENCES (SELECTED)

- Göksel, Aslı, and Celia Kerslake. 2005. *Turkish: A Comprehensive Grammar*. Routledge Comprehensive Grammars. London: Routledge.
- Huguin Mathilde, Barque Lucie, Haas Pauline, Namer Fiammetta et Tribout Delphine. 2022. *Guide d'annotation sémantique des noms construits*, Projet ANR Demonext
- Korkmaz, Zeynep. 2014. *Türkiye Türkçesi Grameri, Şekil Bilgisi*. 4th ed. Ankara, Türkiye: Türk Dil Kurumu Yayınları.
- Körtvélyessy, Livia, Alexandra Bagasheva, and Pavol Štekauer. 2020. *Derivational Networks Across Languages*. Trends in Linguistics / Studies and Monographs. De Gruyter Mouton.
- Namer, Fiammetta, and Nabil Hathout. 2019. 'ParaDis and Démonette: From Theory to Resources for Derivational Paradigms'. In *Second International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2019)*, edited by Žabokrtský Zdeněk, Ševčíková Magda, Litta Eleonora, and Passarotti Marco, 5–14. Proceedings of Derimo - Second International Workshop on Resources and Tools for Derivational Morphology. Prague, Czech Republic: Charles University, Prague. <https://hal.science/hal-02288938>.



**MERCI !**

*Any questions ?*

yagmur.ozturk@edu.univ-fcomte.fr