# Morphological Resources for the Study of Turkish Derived Nouns

Yağmur Öztürk[1], Izabella Thomas[1], Snejana Gadjeva[2]

[1]CRIT, Université de Franche-Comté
[2]CREE, Institut National des Langues et Civilisations Orientales
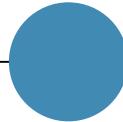
OVERVIEW

Introduction

Turkish Derivational Morphology

NLP Resources in Turkish

Formalisation for Morpheme Description – *Semantürk & DerivBaseTR*

Conclusion – Future Works

"The locations of published resources are **not always stable and/or permanent**. The URLs indicating the location of the resources in papers or on the webpages of the authors or institutions are **not always maintained** and the resources often **disappear** after publication. Although our efforts to reach out to the authors/creators of the resources often yielded positive results, it is desirable to **diminish these barriers** to keep up with the fast-paced research community."

"[A] general area with no or little resources is **semantics**."

Çöltekin et. al (2023)

Construction of Resources for Turkish

Semantürk: ontology of semantic categories

DerivBaseTR: database, describing nominal suffixes in derivational morphology

Resources for Turkish linguistics

Turkish Noun Corpus: corpus of nouns for morphological data extraction
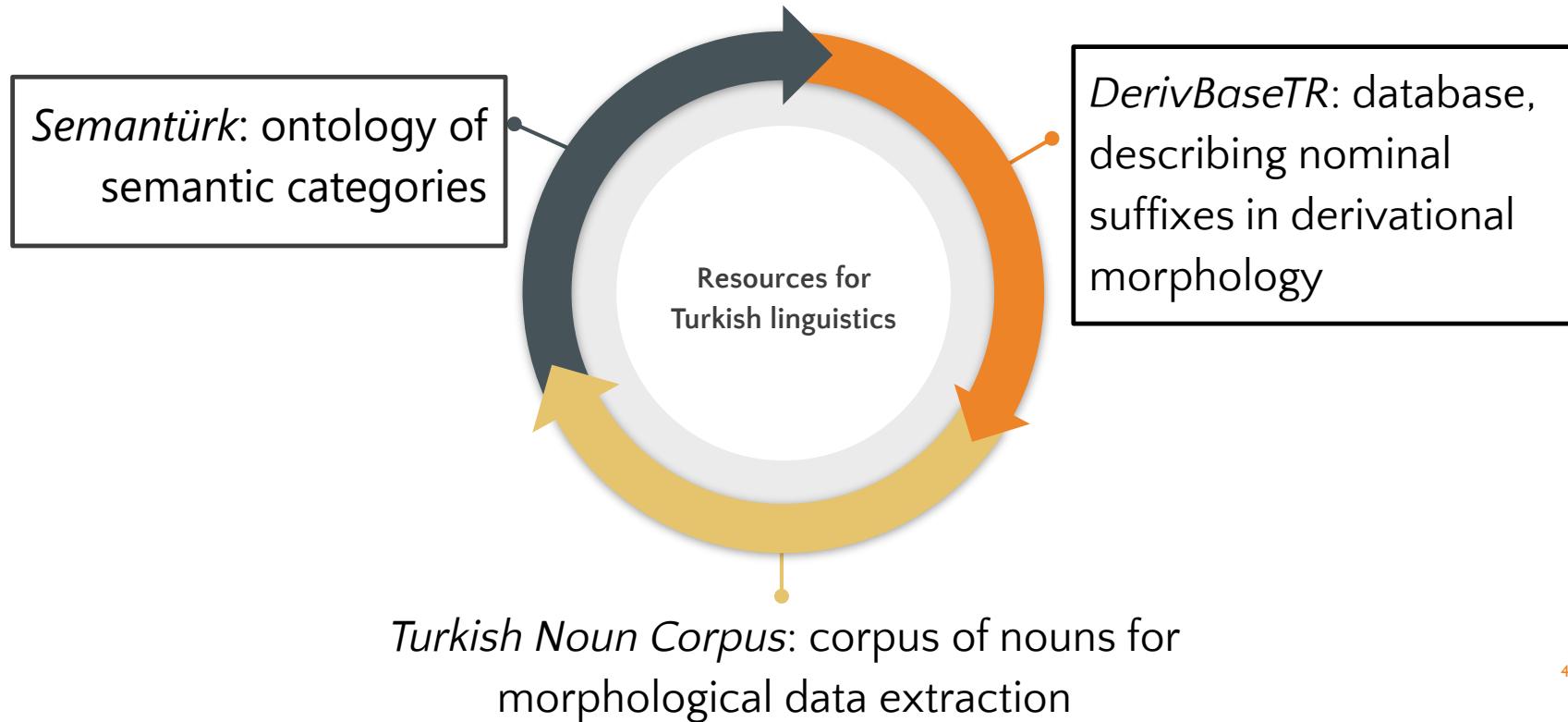
TURKISH DERIVATIONAL MORPHOLOGY

FORMAL, CATEGORIAL & SEMANTIC LEVELS

Formal Level

→ Agglutinative – suffixation

(1)

Turkish	English
Türk	Turk
Türk-çe	Turkish (language)
Fransız-ca	French (language)

(2)

Turkish	English
göz	eye
göz-lük	eyeglasses
göz-lük-çü	optician
göz-lük-çü-lük	opticianry

Formal Level

Simple and complex vowel harmony: –C**A** & –l**İ**k

–A: a, e

–İ: e, i, ü, ö

(1) | **Turkish** | | **English** |
| --- | --- | --- |
| Türk-C**A** | T**ü**rk-ç**e** | Turkish (language) |
| Fransız-C**A** | Frans**ız**-c**a** | French (language) |

(2) | **Turkish** | | **English** |
| --- | --- | --- |
| göz-l**İ**k | g**ö**z-l**ü**k | eyeglasses |
| göz-l**İ**k-C**İ** | g**ö**z-l**ü**k-ç**ü** | optician |

Formal Level

Consonant harmony: –**C**A & – **C**i

$$\boxed{\begin{array}{l} \text{–C: ç, c} \\ \text{–D: t, d} \end{array}}$$

(1)

Turkish		**English**
Türk-**CA**	Tür**k**-**ç**e	Turkish (language)
Fransız-**CA**	Fransı**z**-**c**a	French (language)

(3)

Turkish		**English**
meslek-**DAŞ**	msle**k**-**t**aş	colleague
tür-**DAŞ**	tü**r**-**d**eş	'of the same kind'

Categorial Level

Pluricategoriality of morphemes

(4) **Cat.**   **Turkish**   **English**

    N → **N**   kadın-**ca**   the language of women

    N → **Adj.**   kadın-**ca**   womanlike

    N → **Adv.**   kadın-**ca**   womanly

Semantic Level

Noun–to–Noun
(N–to–N)

Category and semantic correlation

(4)	Cat.	Turkish	English
	N → N	kadın-ca	the language of women
	N → Adj.	kadın-ca	womanlike
	N → Adv.	kadın-ca	womanly

Semantic Level

Polysemy of morphemes

(2) **Turkish**   **English**

    göz   göz   eye

    göz-lİk   gözlük   eyeglasses

    göz-lİk-Cİ   gözlükçü   optician

    göz-lİk-Cİ-lİk   gözlükçülük   opticianry

Semantic Level

Polysemy of morphemes → Distribution

(2) **Turkish**   **English**

    göz   göz   eye

    göz-**lİk**   göz**lük**   eye**glasses**

    göz-**lİk**-Cİ   göz**lük**çü   optician

    göz-**lİk**-Cİ-**lİk**   göz**lük**çü**lük**   optician**ry**

Semantic Level

Synonymy of morphemes

(5) **Turkish**   **English**

    kedi-**Cİk**   kedi-cik   **'the poor little** cat'

    çocuk-**cAğİz**   çocuk-cağız   **'the poor little** kid'

RESOURCES IN TURKISH

NOMINAL MORPHEMES IN NLP TOOLS & LINGUISTIC STUDIES

Morphological Analysers

Available resources:

- TRmorph (Çöltekin, 2010) – Foma FST

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>

Analysis for the word **meslektaş**:

- meslektaş<N>
- meslektaş<N><0><V>
- meslektaş<N><0><V><cpl:pres><3p>
- meslektaş<N><0><V><cpl:pres><3s>

Morphological Analysers

Available resources:

- Trnlp (Bayol, 2018) – Python Library

Input: arkadaşlar *(en. friends)*

Output: arka(isim,sıfat)+daş{İi}[4_26]+lar{Çe}[1_1]

behind(noun,adj)   +daş{NN}   +lar{Infl.}

Morphological Analysers

Available resources:

- Zemberek (Akın and Akın, 2007) – Java Library

- Turkish Morphological Analyzer (Yıldız et al., 2019) – Java Library

# FORMALISED MORPHEME DESCRIPTION
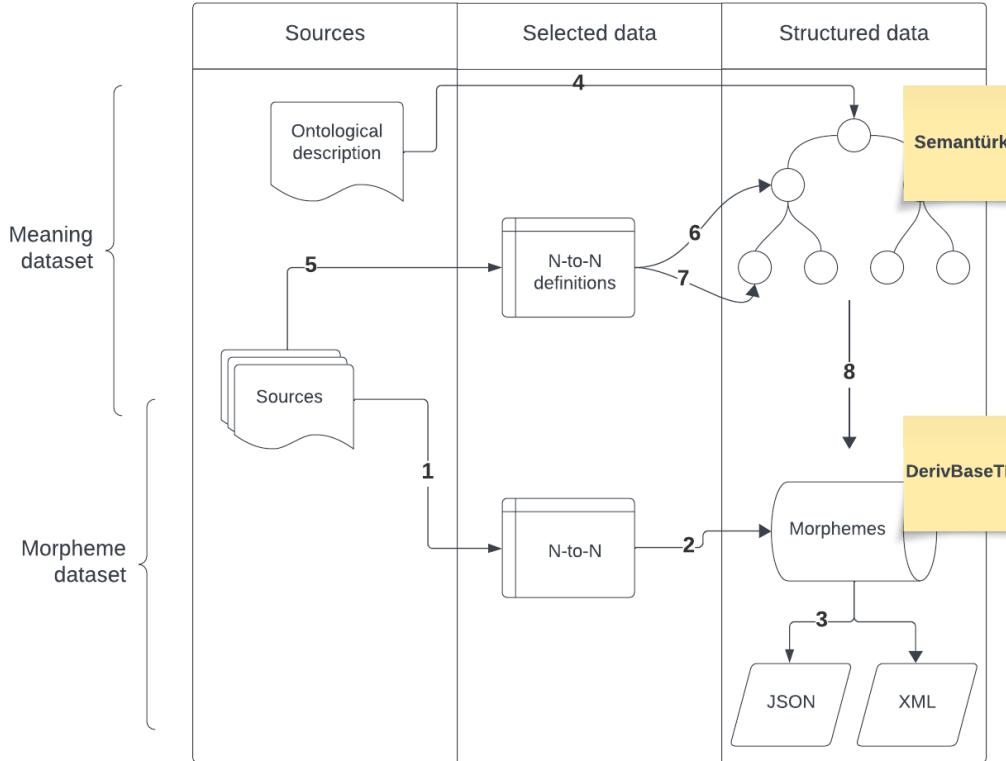
Figure 1: Processing Workflow
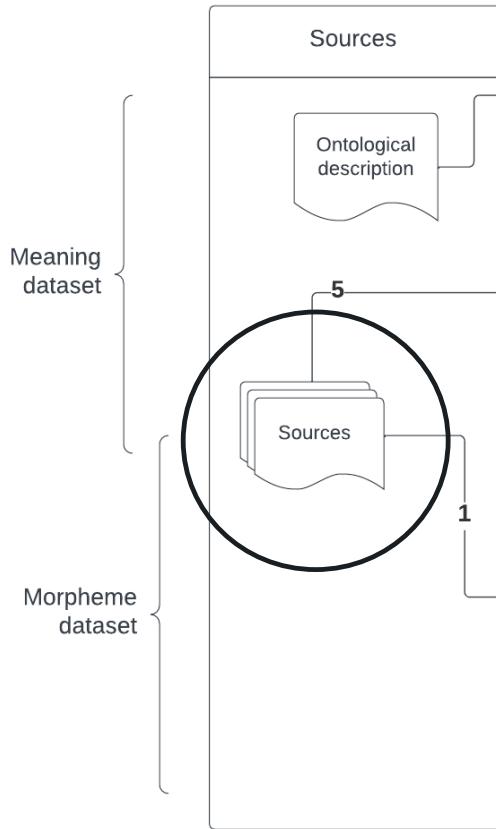
Diverse sources selected:
- Adalı (2004)
- Korkmaz (2014)
- Boz (2015)
- Bazin (1994)
- Göksel and Kerslake (2005)
- Bozdémir (1991)
- Erikan et al. (2008)
- Akçataş and Taşdemir (2020)
- Ozturk (2016)

Linguistic Studies

1. Difference in morpheme categorisation

   – close interaction

   – their ability to function as nominal elements regardless of polycategoriality

   (flexible lexeme ordering in syntax)
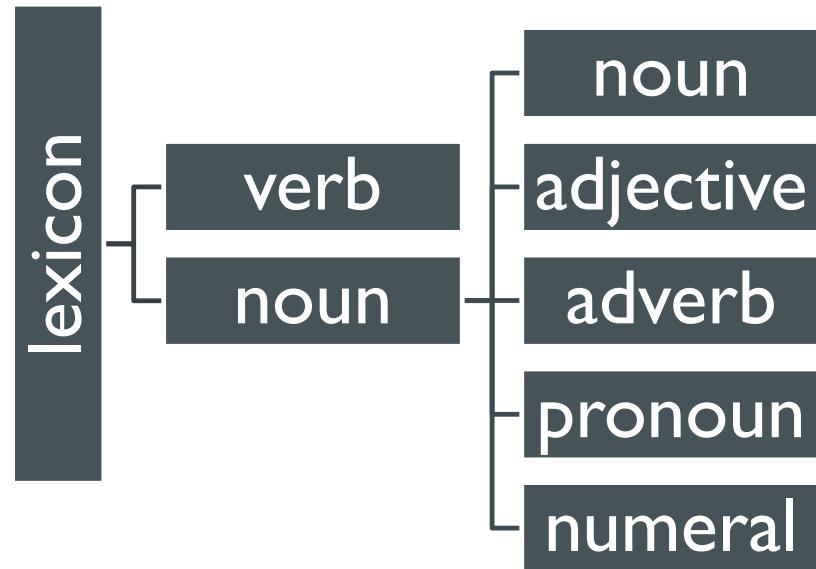
lexicon — verb, noun

verb → noun, adjective

noun → adjective, adverb, pronoun, numeral

Linguistic Studies

1. Difference in morpheme categorisation

a. Categorial flexibility of lexemes is proof of a **functional variation** rather than a categorial variation.

– syntax–dependent

Linguistic Studies

1. Difference in morpheme categorisation

a. Categorial flexibility of lexemes is proof of a **functional variation** rather than a categorial variation.

– syntax–dependent

b. Lexemes **inherently** carry categorial information.

– category can be identified in the lexicon

Linguistic Studies

1. Difference in morpheme categorisation (Görgülü, 2012)

a. Categorial flexibility of lexemes is proof of a **functional variation** rather than a categorial variation.

– syntax-dependent

b. Lexemes **inherently** carry categorial information.

– category can be identified in the lexicon

→ out-of-context

→ online dictionary Türk Dil Kurumu Sözlükleri (*the dictionaries of the Turkish Language Association*): a grammatical category per meaning

Linguistic Studies

2. Incoherence…

      a. throughout single source

– Different set of information ; unsystematic description

      b. across sources

– Formal (e.g. harmony in –cAğIz)
– Categorial (e.g. N–to–N / N–to–Adj. for –sAl)
– Different morpheme entries (e.g. –Cİ, –lİk / –Cİlİk)

…
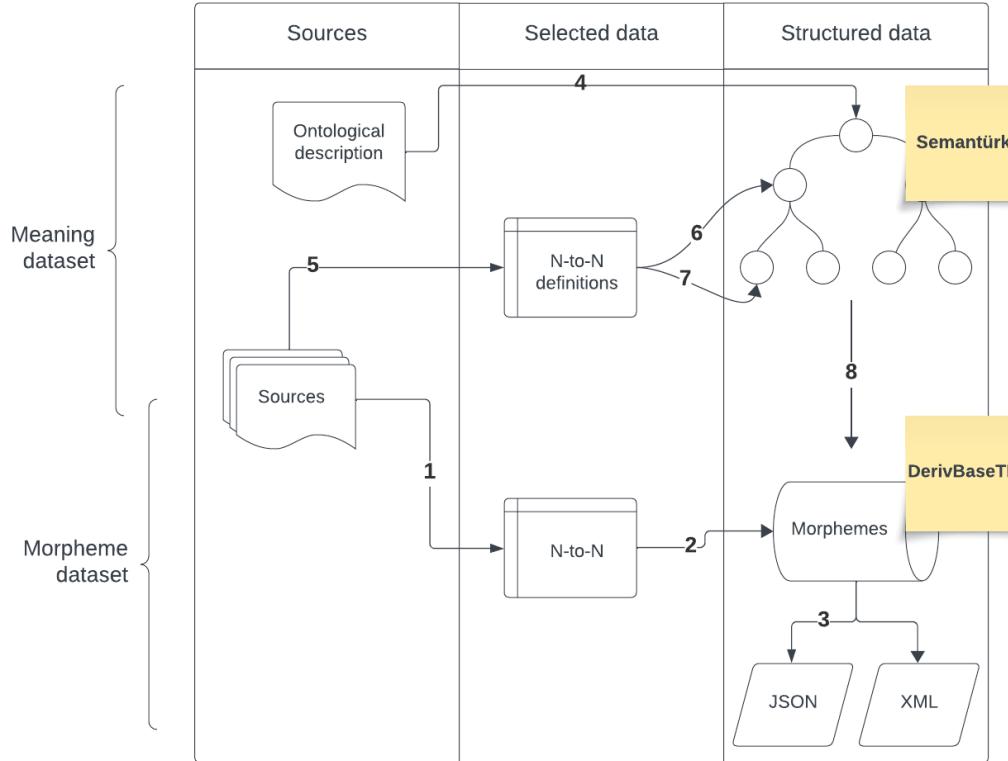
Figure 1: Processing Workflow
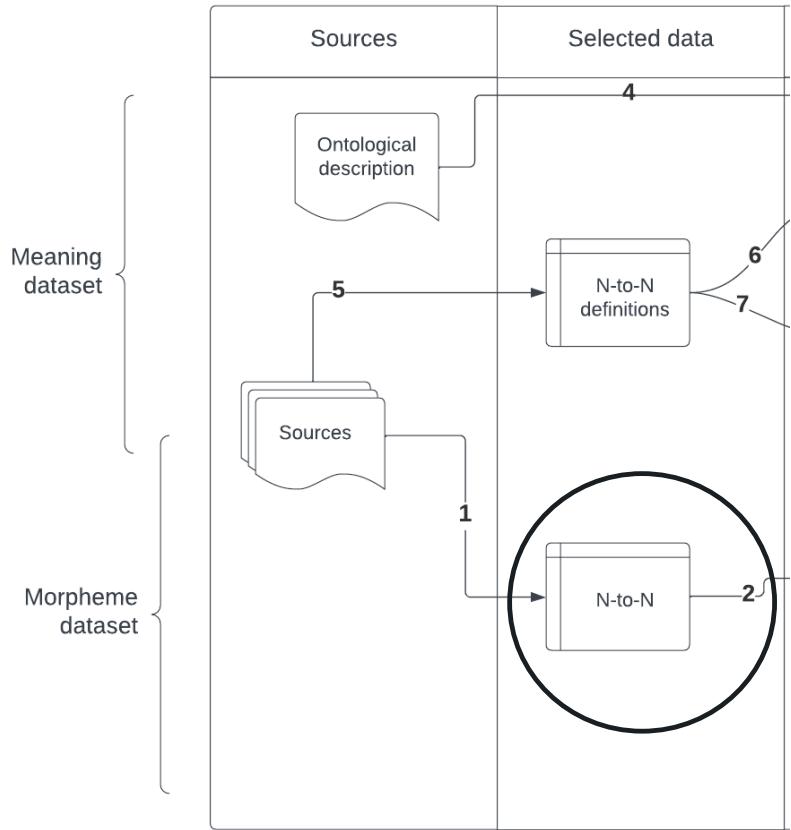
Selection of N–to–N morphemes

Exclusion of Named Entities and dead affixes (lexicalisation)

Borrowed prefixes?

Figure 1: Processing Workflow
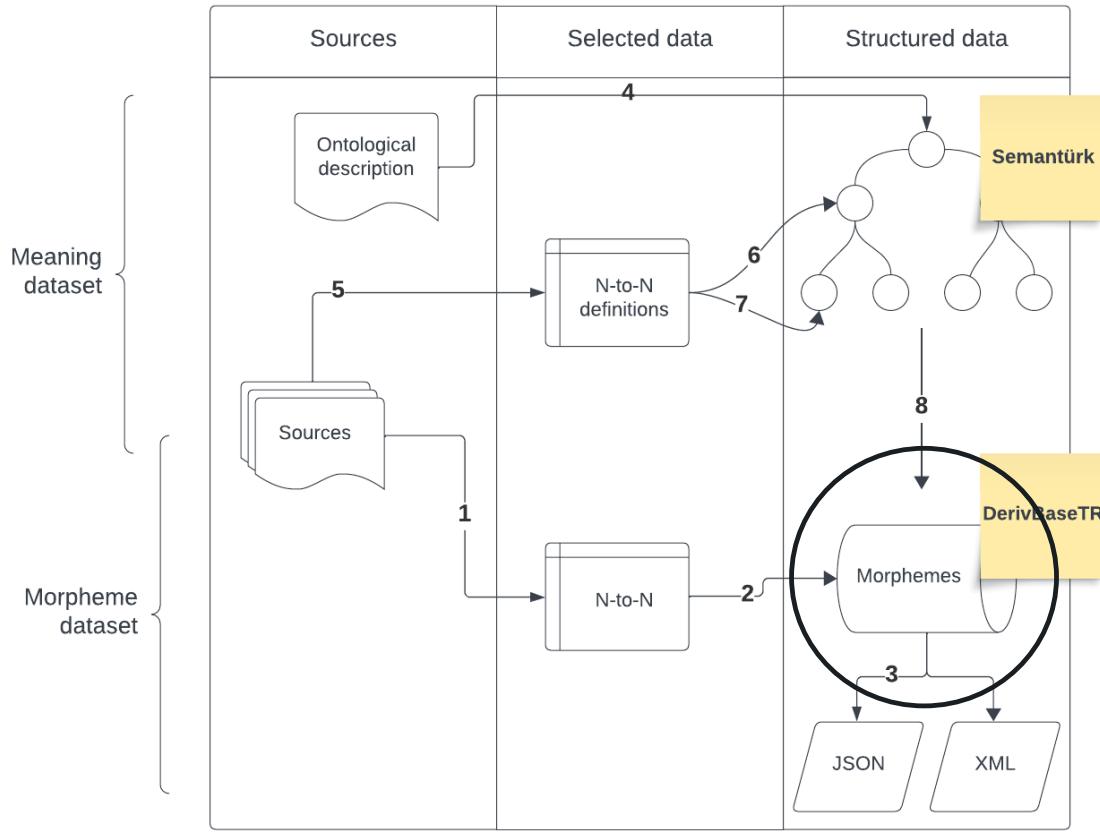
– Formal level

– Categorial level

– Semantic level ?
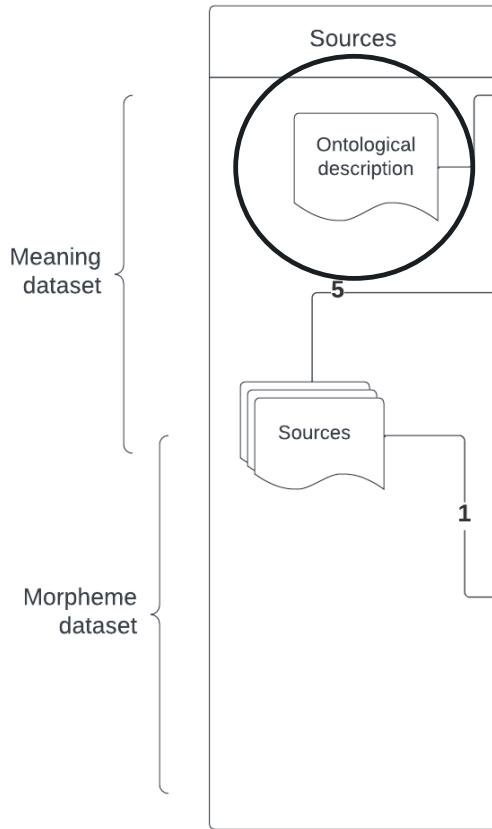
Existing sets of semantic tags :

– Semantic concepts in affixation (Bagasheva, 2017)

– Set of tags from WordNet (Fellbaum, 1998) adapted to Démonnette (Namer. et al, 2019)

→ Annotation scheme by Huguin et al. (2022)
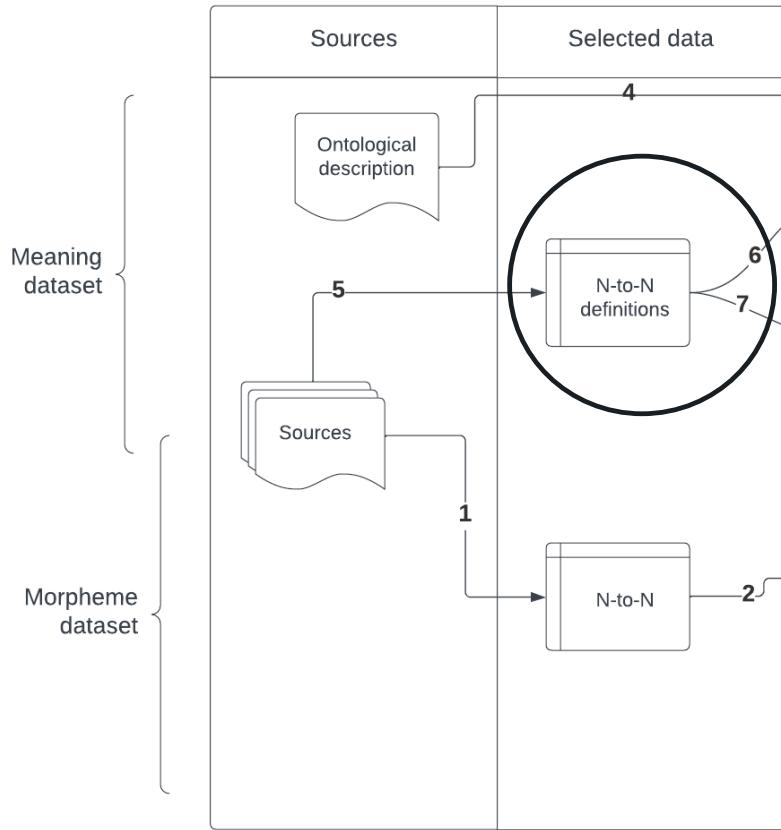
– adaptable to morphosemantics ?
– all of the morphemes semantics ?

→ Collection of morpheme definitions

Semantic Categories

|	
Adalı (2004)	–
Korkmaz (2014)	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**: *adamcağız*, […]
Boz (2015)	**Küçültme, şefkat, acıma sevgi ve alay** – iii. +{cAĞlz} İşlekliğini yitirmiş bir biçimbirimdir: *kız+cağız, kadın+cağız, tavuk+cağız*
Akçataş ve Taşdemir (2020)	+cAğız: Eklendiği kelimeye **"acıma" hissi verir**; *hayvancağız, yavrucağız, çocukcağız…*
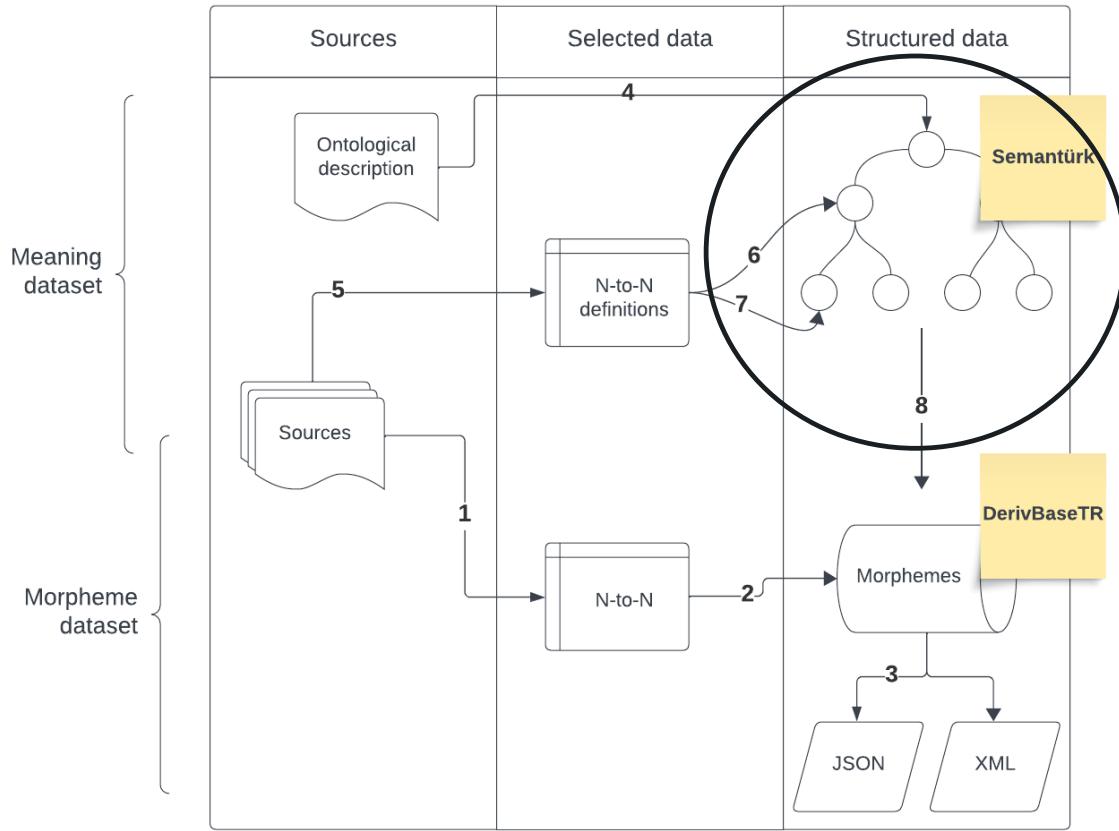
– match with an existing category ?

FEELING
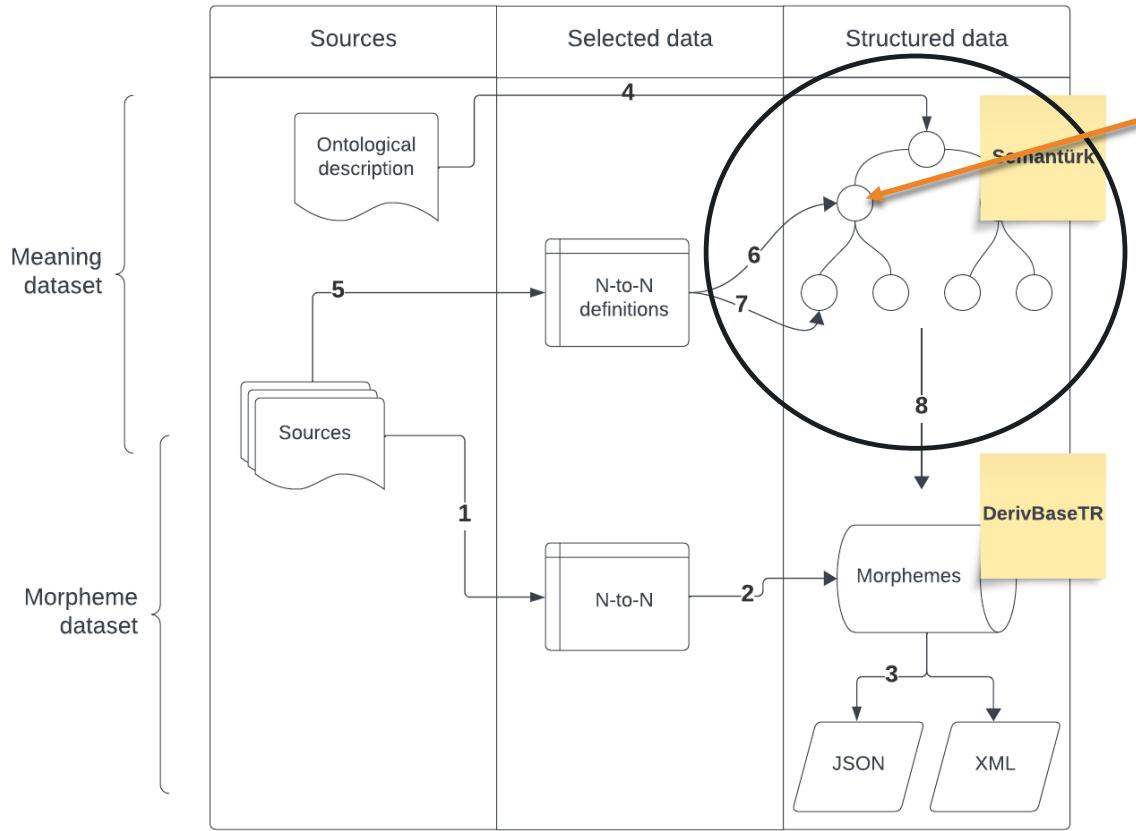
– temporary psychological or physiological state

– the subject is the experiencer of the denoted affect

→ PITY

Figure 1: Processing Workflow

FEELING

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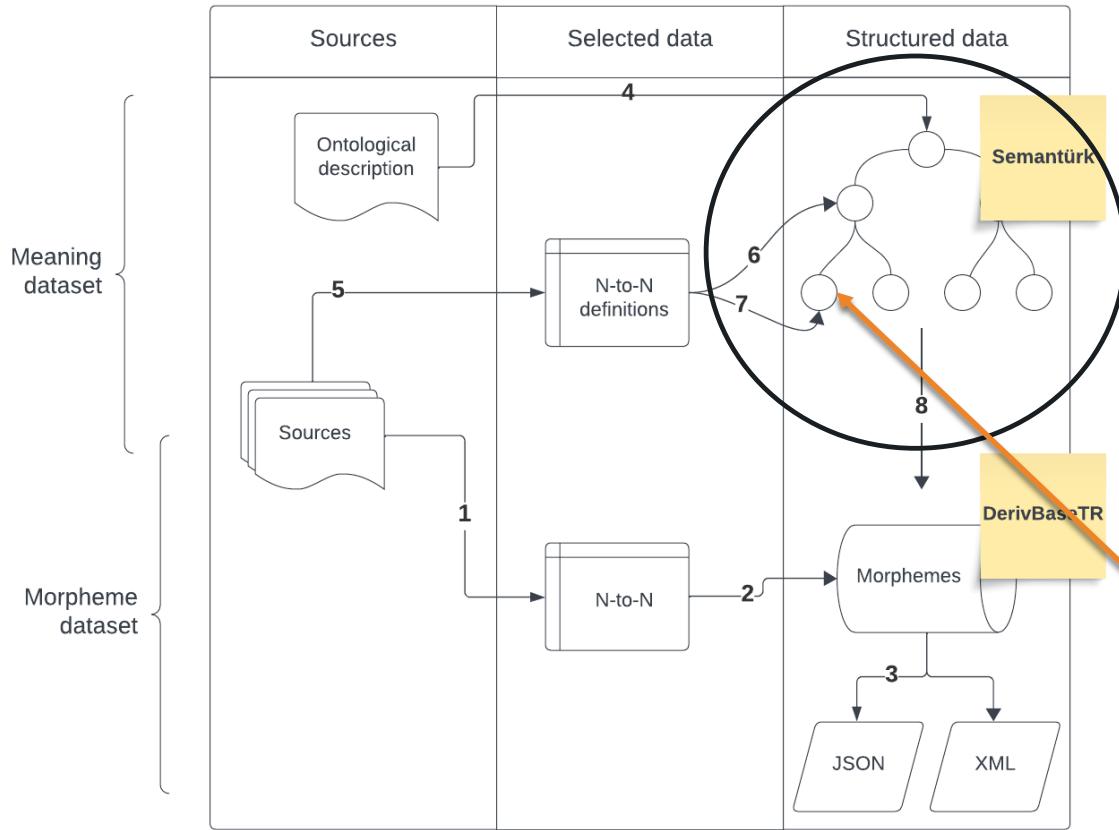
– the subject is the experiencer of the denoted affect

→ PITY
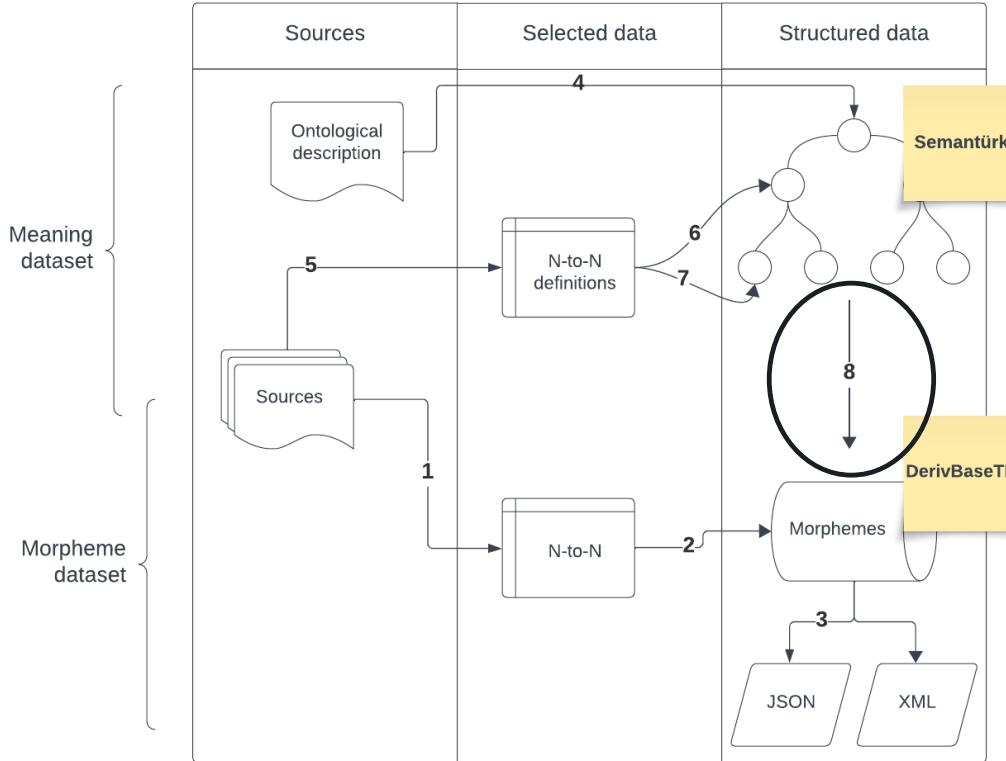
Figure 1: Processing Workflow

Conclusion

DerivBaseTR

— Formalised description:

    morpheme and allomorphs

    base and derived category

    origin

    semantic category(ies)

— Order, select, filter by feature

Conclusion

Semantürk

— Hierarchically organised semantic categories

    Linguistic identifier

    Descriptive identifier

— Usage in other languages (WordNet)

— Description of nominals

> Open Science Framework
> → Interoperable
> → Extendable

To Do or Future Work

Semantürk

❑ Adapt of linguistic tests for each of the semantic categories, comprising the top categories

❑ Evaluate of the semantic categories by Turkish users

DerivBaseTR

❑ Distributive data

❑ Frequency type and/or productivity

REFERENCES

Adalı, Oya. 2004. *Türkiye Tükçesinde 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.

Akın, Ahmet Afsin, and Mehmet Dündar Akın. 2007. 'Zemberek, an Open Source NLP Framework for Turkic Languages'. *Structure* 10 (2007): 1–5.

Bagasheva, A. 2017. « Comparative semantic concepts in affixation ». In: Salvador Valera & Juan Santana (eds.), *Competing Patterns in English Affixation*, 33-65. Peter Lang,

Bayol, Esat Mahmut. 2018. 'Türkçe Doğal Dil İşleme Macerası'. *Türkçe Doğal Dil İşleme – NLP* (blog). 13 August 2018. https://turkceddi.blogspot.com/2018/08/turkce-dogal-dil-isleme-maceras-her.html.

Bazin, Louis. 1994. *Introduction à l'étude Pratique de La Langue Turque*. 3rd ed. Paris: Librairie d'Amérique et d'Orient.

Boz, Erdoğan. 2015. Türkiye Türkçesi Biçimsel ve Anlamsal İşlevli Biçimbilgisi. 4th ed. Ankara, Türkiye: Gazi Kitabevi.

REFERENCES

Bozdémir, Michel. 1991. *Méthode de turc*. Vol. Volume 1. Paris: L'Asiathèque, maison des langues du monde.

Çetinoğlu, Özlem, Orhan Bilgin, and Kemal Oflazer. 2018. 'Turkish Wordnet'. In *Turkish Natural Language Processing*, edited by Kemal Oflazer and Murat Saraçlar, 317–36. Theory and Applications of Natural Language Processing. Cham: Springer International Publishing.

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. https://doi.org/10.1007/s10579-022-09605-4.

Erikan, Catherine, Ayhan Erdal, and Marie Koçoğlu. 2008. 'Leçon 2'. In *Apprenons Le Turc Ensemble / Beraber Türkçe Öğrenelim*, 2nd ed. Vol. Tome 1. Paris: Ataturquie.

Göksel, Aslı, and Celia Kerslake. 2005. Turkish: A Comprehensive Grammar. Routledge Comprehensive Grammars. London: Routledge.

REFERENCES

Görgülü, Emrah. 2012. Semantics of nouns and the specification of number in Turkish. Simon Fraser University.

Huguin Mathilde, Barque Lucie, Haas Pauline, Namer Fiammetta et Tribout Delphine. 2022. *Guide d'annotation sémantique des noms construits*, Projet ANR Demonext, https://hal.archives-ouvertes.fr/hal-03638962.

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.

REFERENCES

Ozturk, Seda. 2016. Création et reconnaissance de néologismes par méthode de suffixation. Université de Franche–Comté, Besançon, France.

Türk Dil Kurumu Sözlükleri (*Les dictionnaires de l'Institut de la langue turque*) [en ligne], disponible sur : https://sozluk.gov.tr/, last accessed: 3rd October 2023.

Yıldız, Olcay Taner, Begüm Avar, and Gökhan Ercan. 2019. 'An Open, Extendible, and Fast Turkish Morphological Analyzer'. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing* (RANLP 2019), 1364–72. Varna, Bulgaria: INCOMA Ltd. https://doi.org/10.26615/978–954–452–056–4_156.

MERCI !

Any questions ?

yagmur.ozturk@edu.univ-fcomte.fr