Building the Turkish FrameNet

Büşra Marşan, Neslihan Kara, Merve Özçelik, Bilge Nas Arıcan, Neslihan Cesur, Aslı Kuzgun, Ezgi Sanıyar, Oğuzhan Kuyrukçu, Olcay Taner Yıldız

Contents

What is FrameNet?

- Introduced by International Computer Science Institute in Berkeley, CA (1997)
- A growing computational lexicography project
- Based on Frame Semantics (Filmore et al., 1976; Filmore, 2006)
- Semantic information on predicate-argument structure

What is FrameNet? / Frame

- A schematic representation of an event, state or relationship
- Constituted of individual lemmas (also known as Lexical Units) and frame elements (such as the agent, theme, instrument, duration, manner, direction etc.)
- Lexical Units, or lemmas, are linked to a frame through **a single sense.**

What is FrameNet? / Frame

roast (v) (1)

to criticise harshly

roast (v) (2)

to cook by exposing to dry heat (as in an oven or before a fire) or by surrounding with hot embers, sand, or stones

What is FrameNet? / Frame

Apply_heat

Definition:

A Cook applies heat to Food, where the Temperature_setting of the heat and Duration of application may be specified. A Heating_instrument, generally indicated by a locative phrase, may also be expressed. Some cooking methods involve the use of a Medium (e.g. milk or water) by which heat is transferred to the Food. A less semantically prominent Food or Cook is marked Co-participant.

Sally FRIED an egg in butter.

Sally FRIED an egg in a teflon pan.

Ellen FRIED the eggs with chopped tomatoes and garlic.

This frame differs from Cooking_creation in focusing on the process of handling the ingredients, rather than the edible entity that results from the process.

https://framenet.icsi.berkeley.edu/fndrupal/frameIndex

Our Motivation

- Creating a comprehensive Turkish FrameNet
- Ensuring a certain level of correspondence between English and Turkish FrameNets
- Making Turkish FrameNet compatible with other resources in the language:
  - KeNet (Turkish Wordnet)
  - TRopBank (Turkish PropBank)

A "Compatible" FrameNet

- Using more than one resource in an NLP task is often challenging:
  - Different teams create resources like WordNet, PropBank or FrameNet
- In Turkish, the same team created all these resources, using
  - Same dataset
  - Same IDs
  - Same annotated sample sentences

Towards the Turkish FrameNet / Methodology

- Two main strategies:
  - Frame-by-frame approach: *Preferred by the vast majority*
  - Lemma-by-lemma approach: *German FrameNet (SALSA), Japanese FrameNet*
- Both offer a set of advantages and challenges.

Towards the Turkish FrameNet / Methodology

- Frame-by-frame approach
  - Coherency within the frames
  - Developing relationships between frames are easier
- Lemma-by-lemma approach
  - Unveils the full semantic range of a given lemma
  - More comprehensive analysis of LUs
  - More "biased" lexicon
  - Developing frames and relationships between frames are more difficult

- Our choice was based on data and time constraints.
- A hybrid approach (Candito et. al, 2014, French FrameNet)
  - Efficient
  - Fast
  - "Enough" coverage
- Detected 8 major domains

Towards the Turkish FrameNet / Domain Selection Process

- We examined Turkish wordnet and detected 8 major domains with most entries and relatively higher frequency:
  - Activity
  - Cause
  - Change
  - Motion
  - Cognition
  - Perception
  - Judgement
  - Commerce

Towards the Turkish FrameNet / Domains

- Frames from English FrameNet were adopted when possible.
- New frames were created when needed.
- Each annotator focused on one domain.
- Second step will involve a lemma-by-lemma approach.

Towards the Turkish FrameNet / Annotator Team

- English FrameNet: 3 different teams.
  - Vanguards
  - Annotators
  - Rearguards
- We opted for a different strategy.

Towards the Turkish FrameNet / Annotator Team

- We opted for a different strategy:
  - Four teams of two.
  - Each annotator was given a domain.
  - Team members were always in touch. Also weekly meetings were held to cross-check annotations of different teams.

- Annotators created frames within their domain by translating and adopting related frames from English FrameNet.
- Then they had to extract lexical units from TRopBank and KeNet, annotate their frame elements, write sample sentences and annotate these sentences.

Towards the Turkish FrameNet / Inter Annotator Agr.

- Weekly meetings.
- After annotation process was done, a team of three carefully assessed all annotations.
- Agreement metrics:
  - Coherence with English FrameNet (if possible)
  - Overlap
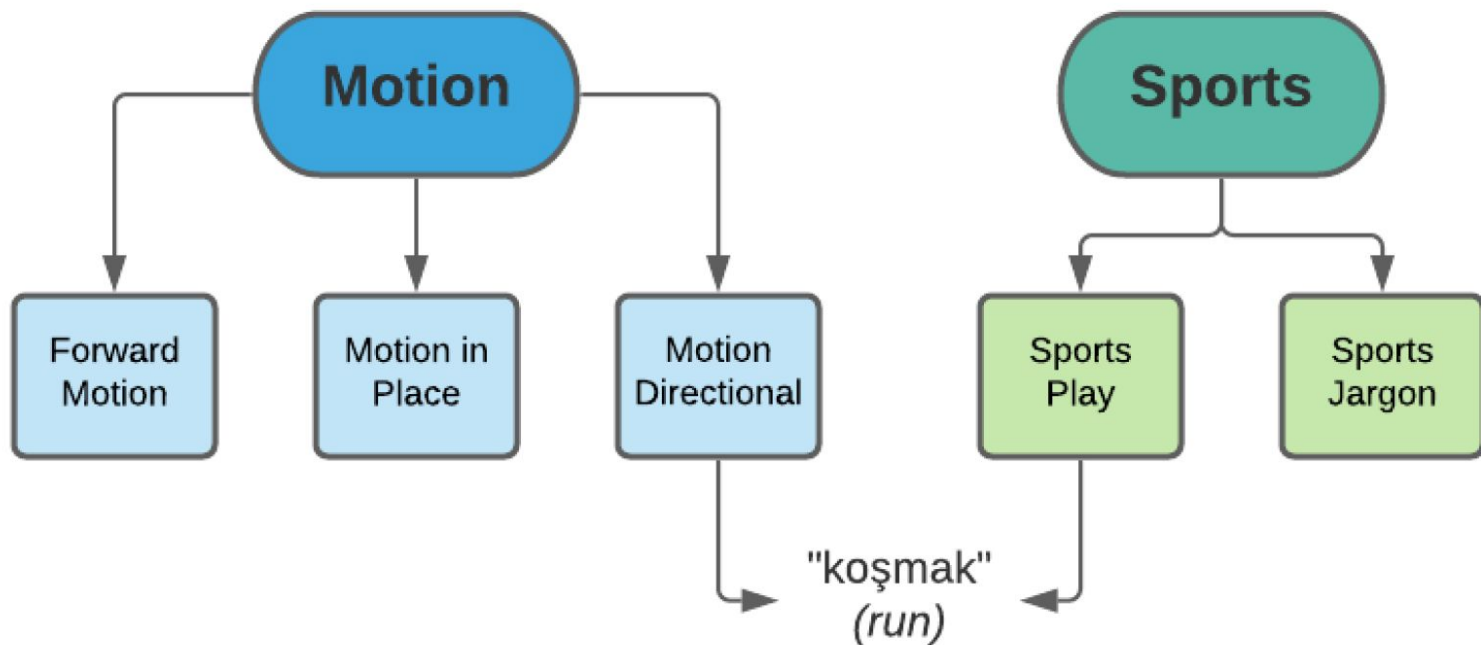
Towards the Turkish FrameNet / LU & Synset

- Compatibility with other resources.
- Lexical Units were KeNet (Turkish Wordnet) synsets.
- Members of the synsets:
  - Share a definition,
  - Have same syntactic properties (Theta role assignment, case assignment etc),
  - Can be used interchangeably without any meaning loss.

Towards the Turkish FrameNet / LU & Synset

Frame	Lexical Unit ID	Synset	Definition
Apply_Heat	TUR10-0354260	ısıtmak, sıcaklaştırmak	Sıcak duruma getirmek
Apply_Heat	TUR10-1154650	tava getirmek	Gereği kadar ısıtmak
Apply_Heat	TUR10-0810920	ütmek	Taze buğday veya mısırı ateşe tutup pişirmek
Apply_Heat	TUR10-0810910	ütmek	Bir şeyi, tüylerini yakmak için alevden geçirmek

Towards the Turkish FrameNet / Annotation Interfaces

Towards the Turkish FrameNet / Data

- Predicates were taken from KeNet and TRopBank.
- Sample sentences were taken from TDK's dictionary when possible. Otherwise created by the team of annotators.
- Also dependency trees of sample sentences were created.

Towards the Turkish FrameNet / Data

Frame	Definition	Lexical Unit Id	Lexical Unit Synset	Lexical Unit Definition	Frame Elements
Attempt	An Agent attempts to achieve a Goal. The Outcome may also be mentioned explicitly.	TUR10-0192570	denemek	Bir işe, başarmak amacıyla başlamak, girişimde bulunmak, teşebbüs etmek	Agent, Goal, Circumstances, Effort
		TUR10-1160410	teşebbüse geçmek	bir işi yapmak için davranmak, girişmek	Agent, Goal, Circumstances, Effort, Manners
		TUR10-1160420	teşebbüs etmek	girişmek, el atmak	Agent, Goal, Circumstances, Effort, Manners
		TUR10-0479350	koyulmak	Girişmek, başlamak, teşebbüs etmek	Agent, Goal, Manners
		TUR10-1032280	girişimde bulunmak	davranmak, teşebbüs etmek	Agent, Goal, Circumstances, Effort, Manners
		TUR10-1183220	yerini yapmak	bir şey elde etmek amacıyla girişimde bulunmak	Agent, Goal, Circumstances, Effort, Manners
		TUR10-0298660	girişmek	Kalkışmak	Agent, Goal, Manners

Turkish FrameNet, Open & Available / GitHub

You can find FrameNet data on GitHub now, and on Global Framenet Project databse soon!

StarlangSoftware

Starlang Yazılım Danışmanlık Sanayi ve Ticaret Limited Şirketi

⊙ Istanbul / Turkey 🔗 http://www.starlangyazilim.com ✉ info@starlangyazilim.com

▯ Repositories 163 📦 Packages 👤 People 13 👥 Teams 1 ▥ Projects

https://github.com/StarlangSoftware/TurkishFrameNet
https://github.com/StarlangSoftware/TurkishFrameNet-Py
https://github.com/StarlangSoftware/TurkishFrameNet-Cy
https://github.com/StarlangSoftware/TurkishFrameNet-C#
https://github.com/StarlangSoftware/TurkishFrameNet-CPP

Turkish FrameNet, Open & Available / GitHub

TurkishFrameNet-CPP

● C++   ⑂0   ☆0   ⓘ0   ⑂0   Updated 17 days ago

TurkishFrameNet-Cy

● Python   ⑂0   ☆0   ⓘ0   ⑂0   Updated 2 days ago

TurkishFrameNet-CS

● C#   ⑂0   ☆0   ⓘ0   ⑂0   Updated 19 days ago

TurkishFrameNet-Swift

● Swift   ⑂0   ☆0   ⓘ0   ⑂0   Updated 19 days ago

TurkishFrameNet-Py

● Python   ⑂0   ☆0   ⓘ0   ⑂0   Updated 2 days ago

TurkishFrameNet

● Java   ⑂0   ☆0   ⓘ0   ⑂0   Updated 19 days ago

https://github.com/StarlangSoftware/TurkishFrameNet

https://github.com/StarlangSoftware/TurkishFrameNet-Py

https://github.com/StarlangSoftware/TurkishFrameNet-Cy

https://github.com/StarlangSoftware/TurkishFrameNet-C#

https://github.com/StarlangSoftware/TurkishFrameNet-CPP

Turkish FrameNet / Statistics

Table 2: Statistics

Total Frames	139
Unique Frames	16
Synsets (LUs)	2561
Individual Predicates	4080
Frame Elements	203

Table 3: A comparison with initial versions of other FrameNets

Language	Frames	LUs
French	98	662
Chinese	322	3947
Swedish	51	2300

Frames / Challenges

- Typological differences between Turkish and English caused some challenges.
- Turkish has more:
  - Unaccusative verbs,
  - Lexicalized, figurative multi-word predicates.

dinmek (v) *"to stop"*   vs.   **dondurmak** (v) *"to freeze"*

- Different valency and argument structure, but both confirm to the definition of Activity_Paused_State frame.

Frames / Splitted Frames

- Frugality frame was split:
  - Frugality (20 LUs)
  - Frugality_Time (6 LUs)
  - Frugality_Waste (16 LUs)
  - Frugality_Money (18 LUs)

Frames / Splitted Frames

Frugality		TUR10-1053470	tüketmek ifna etmek istihlak etmek yiyip bitirmek yoğaltmak	Kullanarak, harcayarak yok etmek, bitirmek
Frugality		TUR10-0889330	dibine darı ekmek	bir şeyi sonuna kadar tüketmek, bitirmek
Frugality		TUR10-0716370	süpürmek	Tüketmek; bitirmek
Frugality		TUR10-0899090	har vurup harman savurmak	Düşüncesizce ve hesapsızca harcamak, bol bol harcayıp tüketmek
Frugality		TUR10-1005890	dibini bulmak	içindekini tüketmek
Frugality		TUR10-0888390	dağarcıkta bir şey kalmamak	her şeyi tüketmek, bitirmek
Frugality		TUR10-1207070	çarçur etmek	Gereksiz yere harcayıp tüketmek
Frugality	The words of this frame describe how a Resource_controller spends money or other Resources for a particular purpose . They can be applied to the Resource_controller or to their Behavior in particular instances. These words indicate a judgment of how conservatively the Resource_controller guards their Resource.	TUR10-0934310	altından girip üstünden çıkmak	malı, parayı düşüncesizce harcayıp tüketmek
Frugality		TUR10-0899470	havaya savurmak	gereksiz yere harcamak
Frugality		TUR10-1138240	sokağa atmak	para, eşya vb.ni boş yere harcamak
Frugality		TUR10-0489340	kullanmak	Harcamak; sarf etmek
Frugality		TUR10-0847730	yemek	Harcamak; tüketmek; bitirmek
Frugality		TUR10-1143630	suya salmak	boşuna harcamak
Frugality		TUR10-0220340	harcamak dökmek sarf etmek vermek	Bir şey yapmak için kullanmak, tüketmek
Frugality		TUR10-0250870	eritmek	Harcayıp tüketmek
Frugality		TUR10-0939180	araya vermek	yararsız bir işe harcamak
Frugality		TUR10-0327650	harcamak	Bir iş görmek veya bir şey satın almak için parayı elden çıkarmak, sarf etmek
Frugality		TUR10-1061220	istimal etmek kullanıvermek sarf etmek sarfetmek tutmak yer vermek	Kullanmak
Frugality		TUR10-0288560	geçmek ezmek harcama yapmak yatırmak harcetmek	Harcamak
Frugality		TUR10-0880650	çıkmak	Harcamak zorunda kalmak

Frames / Splitted Frames

Frugality_Time		TUR10-1172380	vakit geçirmek oyalanmak kalmak siftinmek	Boş boş vakit harcamak
Frugality_Time	The words of this frame describe how a Resource_controller spends time for a particular purpose . They can be applied to the Resource_controller or to their Behavior in particular instances.	TUR10-1046810	havyar kesmek	çalışmadan vakit geçirmek, vakti boşa harcamak
Frugality_Time		TUR10-1209900	zaman geçirmek	Oyalanmak; vakit geçirmek
Frugality_Time		TUR10-0287930	geçirmek	Zaman harcamak
Frugality_Time		TUR10-0628080	pineklemek	Boşa zaman harcamak
Frugality_Time		TUR10-0914640	vakit saat aramamak	zamana hiç aldırmamak

Frames / Splitted Frames

Frugality_Waste		TUR10-1117160	savurmak parayı denize atmak israf etmek saçıp savurmak	Boşuna ve çok miktarda para harcamak, parayı düşüncesizce ve boşuna harcamak
Frugality_Waste		TUR10-1047730	ziyan etmek heder etmek heba etmek murdar etmek mahvetmek	Boşa gitmesine sebep olmak, yersiz, boş yere harcamak
Frugality_Waste		TUR10-1181620	yele vermek	Savurmak; boşuna harcamak
Frugality_Waste		TUR10-0627200	piç etmek	Boş geçirmek, boşa harcamak
Frugality_Waste	The words of this frame describe how a Resource_controll er spends (their) resources in a reckless, wasteful way. They can be applied to the Resource_controll er or to their Behavior in particular instances.	TUR10-1125120	saçılıp dökülmek	gereğinden veya kaldırabileceğinden çok harcamak
Frugality_Waste		TUR10-0896430	gâvur etmek	boşuna harcamak, yerinde harcamamış olmak, işe yaramaz duruma getirmek
Frugality_Waste		TUR10-1116870	para saçmak	gereğinden çok para harcamak
Frugality_Waste		TUR10-0899470	havaya savurmak	gereksiz yere harcamak
Frugality_Waste		TUR10-1138240	sokağa atmak	para, eşya vb.ni boş yere harcamak
Frugality_Waste		TUR10-0934310	altından girip üstünden çıkmak	malı, parayı düşüncesizce harcayıp tüketmek
Frugality_Waste		TUR10-0939180	araya vermek	yararsız bir işe harcamak
Frugality_Waste		TUR10-1117100	para yemek	gereksiz olarak çok para harcamak
Frugality_Waste		TUR10-0899090	har vurup harman savurmak	Düşüncesizce ve hesapsızca harcamak, bol bol harcayıp tüketmek
Frugality_Waste		TUR10-0239040	ekmek	Parayı boşuna harcamak, ziyan etmek
Frugality_Waste		TUR10-1116910	parasını sokağa atmak	değeri olmayan bir mala para vermek
Frugality_Waste		TUR10-1060590	israfa kaçmak	gereksiz yere aşırı harcamalarda bulunmak

Frames / Splitted Frames

Frugality_Money		TUR10-1242660	harcamak	Parayı elden çıkarmak
Frugality_Money		TUR10-1116870	para saçmak	gereğinden çok para harcamak
Frugality_Money		TUR10-0327650	harcamak	Bir iş görmek veya bir şey satın almak için parayı elden çıkarmak, sarf etmek
Frugality_Money		TUR10-0910400	masraf etmek	para harcamak
Frugality_Money		TUR10-1095990	masrafa girmek	bir iş veya yapım için çok para harcamak
Frugality_Money	The words of this frame describe how a Resource_controller spends money for a particular purpose and often, in a particular way. They can be applied to the Resource_controller or to their Behavior in particular instances.	TUR10-1116450	para dökmek para akıtmak	çok para harcamak
Frugality_Money		TUR10-1116980	paraya kıymak	gereken yerde para harcamaktan kaçınmamak
Frugality_Money		TUR10-1117010	paraya para dememek paraya pul dememek	bol para harcamak
Frugality_Money		TUR10-1116550	paradan çıkmak masraftan çıkmak	beklenmedik bir sırada para harcama durumunda kalmak
Frugality_Money		TUR10-0847760	yemek	Başkasının parasını harcamak
Frugality_Money		TUR10-0907220	kesenin ağzını açmak	bol para harcamaya başlamak
Frugality_Money		TUR10-1167220	ulufe dağıtmak ulufe vermek	yerli yersiz bol keseden para harcamak
Frugality_Money		TUR10-0711430	sulamak	Para ödemek, vermek, harcamak
Frugality_Money		TUR10-0705580	sökülmek	Parayı istemeyerek vermek, harcamak
Frugality_Money		TUR10-0349020	hovardalık etmek	Zevki için bol para harcamak
Frugality_Money		TUR10-1117080	para yedirmek	gereksiz olarak başkasına çok para harcamak
Frugality_Money		TUR10-1042620	hacıağalık etmek	gereksiz yere, gösteriş için bol para harcamak
Frugality_Money		TUR10-1117100	para yemek	gereksiz olarak çok para harcamak

- New frames were created:
    - Deprive
    - Cognitive Change
    - Depict
    - Expulse
    - *(12 more)*

Discussion

- A satisfactory first step towards creating a comprehensive and coherent Turkish FrameNet.
- Correspondence to English FrameNet makes it possible to introduce cross-correspondence between Turkish and various other FrameNets that use same or similar frames as English FrameNet.
- Streamlined databanks and corpora: KeNet, TRopBank and FrameNet

Further Studies

- A bottom-up approach to extend the coverage of FrameNet.
- Using KeNet's own hierarchy as a guideline to create new frames.

Thank you for listening!

If you'd like to get in touch, feel free to e-mail us!

busra@starlangyazilim.com

neslihan@starlangyazilim.com

info@starlangyazilim.com