

FrameForm: An Open-source Annotation Interface for FrameNet

Büşra Marşan

Starlang Yazılım Danışmanlık

busra@starlangyazilim.com

Olca Taner Yıldız

Özyeğin University

olcay.yildiz@ozyegin.edu.tr

Abstract

In this paper, we introduce FrameForm¹, an open-source annotation tool designed to accommodate predicate annotations based on Frame Semantics (Fillmore et al., 1976). FrameForm is a user-friendly tool for creating, annotating and maintaining computational lexicography projects like FrameNet and has been used while building the Turkish FrameNet (Marşan et al., 2021). Responsive and open-source, FrameForm can be easily modified to answer the annotation needs of a wide range of different languages.

1 Introduction

FrameNet (Lowe, 1997; Baker et al., 1998; Fillmore and Atkins, 1998; Johnson et al., 2001) is a growing NLP resource developed by the International Computer Science Institute in Berkeley, California. Having its theoretical background in Fillmore’s Frame Semantics notion (Fillmore et al., 1976), FrameNet is a coherent and exhaustive computational lexicography that provides in-depth semantic information regarding the argument structure and thematic relations of a predicate.

In FrameNet, predicates are annotated into their respective frames. A frame refers to a schematic representation that brings lemmas together based on their semantic properties and syntactic features (for a more detailed definition and discussion of the frame notion, see (Fillmore et al., 1976)). For instance, Motion frame brings frames that denote a motion between two points or on a path, and it has the following definition²:

Some entity (Theme) starts out in one place (Source) and ends up in some other place (Goal), having covered some

space between the two (Path). Alternatively, the Area or Direction in which the Theme moves or the Distance of the movement may be mentioned.

Predicates (Lexical Units or LUs, as called in FrameNet) that fit the definition above are annotated to this frame. Here we must point out that each sense of a Lexical Unit pertains to a different frame. For instance, “blow” (as an intransitive verb) has several meanings³:

1. to move or be carried by or as if by wind
2. erupt, explode
3. to send forth a current of air or other gas

With its first meaning, “blow” is annotated to Move frame.

Following the framework put forward by English FrameNet team, many other researchers re-created this resource in different languages. In order to ease building process and streamline the maintenance of FrameNet resources, we developed an open-source annotation interface called FrameForm⁴.

We will discuss the creation process of FrameForm in Section 2 and introduce its features in Section 3. Finally, we will offer a brief discussion regarding the future work in Section 4.

2 Developing the FrameForm

Development process of FrameForm is closely tied to the creation process of Turkish FrameNet (Marşan et al., 2021): While gathering data for Turkish FrameNet, our team needed an easy-to-use tool that allowed frame annotation, semantic annotation and morphologic analysis. First we did a thorough research to see what software and tools

¹<https://github.com/StarlangSoftware/SemanticRoleLabeling>

²<https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>

³<https://www.merriam-webster.com/dictionary/blow>

⁴<https://github.com/StarlangSoftware/SemanticRoleLabeling>

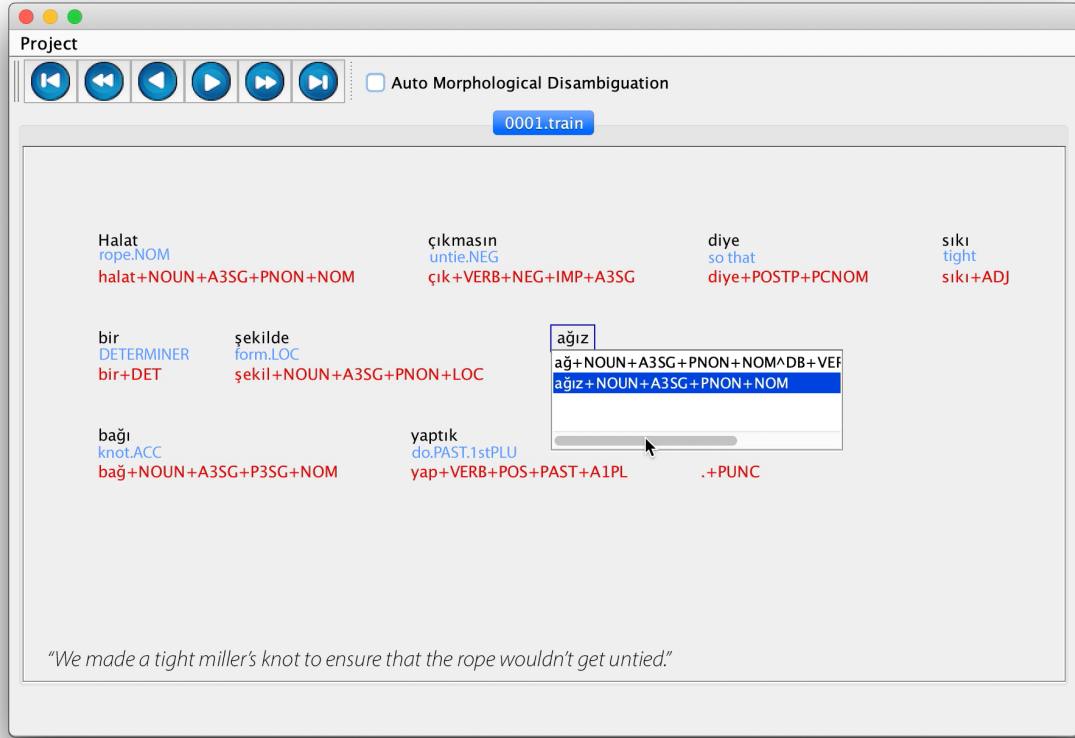


Figure 1: Morphologic analysis interface of FrameForm

were used for building other FrameNets. In their articles covering the process of building and/or expanding their FrameNets, many teams don't mention the tools or interface they use, that is why we were able to find only few resources regarding the annotation of FrameNets in various languages:

- FrameNet Brasil team uses a web-based annotation tool called FrameNet Brasil WebTool (Matos and Torrent, 2016). The same tool is also used for Global FrameNet annotations. It allows the users to create language specific tags to accommodate typological features of different languages but it does not allow an in-depth morphological analysis. That is why our team was unable to use FrameNet Brasil WebTool.
- The team behind German FrameNet SALSA uses two main tools for annotation: SALTO (Burchardt et al., 2006) and FrameNet Transformer (Ruppenhofer et al., 2010). Although very practical, these tools fell short of satisfying our needs regarding morphological analysis and semantic annotation.

- Swedish FrameNet team uses Karp, “the open lexical infrastructure of Sprakbanken (the Swedish Language Bank)” (Borin et al., 2013), which cannot be used for annotating other languages.
- Spanish FrameNet team uses the same annotation software as Berkeley team (Fillmore et al., 2002), which, again, does not allow us to do a semantic annotation and morphological analysis as detailed as we desire.

After our thorough research, we found ourselves in a position where we had to develop our own annotation interface. Thus, we created FrameForm. It is written in Java and can be found on GitHub. Since it is an open-source program, it is possible to change or further develop FrameForm freely. That is why we believe that it can be easily integrated into many other FrameNet projects in different languages.

Thus far, mostly Indo-European languages followed suit with the English FrameNet. These languages are relatively poorer in morphology compared to agglutinative languages like Turkish. That

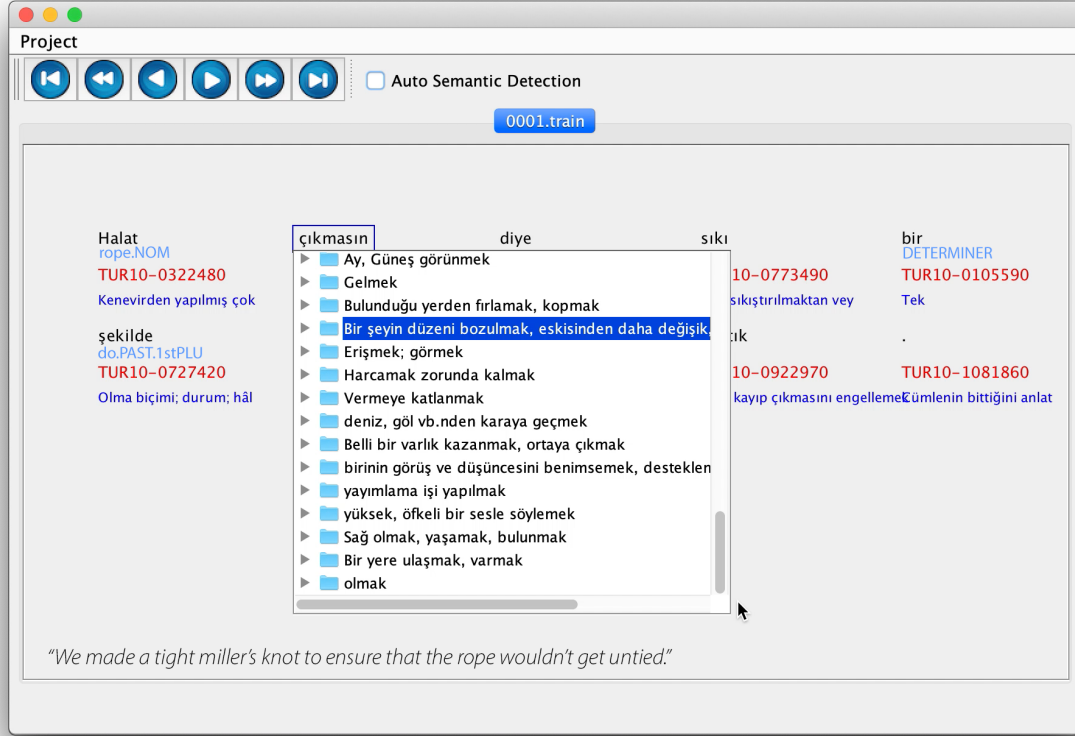


Figure 2: Semantic annotation interface of FrameForm

is why the annotation tools we discussed above do not offer a morphologic analyser component which is essential for morphologically richer languages. Our annotation tool, FrameForm allows adding a new morphological analyser and introducing a new dictionary and/or WordNet. That is why different languages can utilise FrameForm for their annotation processes simply by making some minor adjustments in the back-end.

3 Features

FrameForm saves every annotation pertaining to a Lexical Unit in a single file: Morphologic analysis, predicate analysis (shows which word or group of words is the predicate), semantic analysis (maps the related meaning to the word), frame information and frame elements. This way, one can find all the necessary information regarding a Lexical unit with one click.

The annotation process starts with the morphologic analysis of the sample sentence (see Figure 1). For this analysis, we incorporated our own morphological analysis library for Turkish, which can

be accessed freely on GitHub⁵. Using this library, FrameForm offers an automatic morphologic analysis to speed up the process. The annotator can change auto-generated annotation if it is not correct.

Using the Starlang Turkish Morphological Analysis library, FrameForm processes roots and suffixes separately. It chooses the longest possible root (including derivational suffixes but excluding inflectional suffixes). If the longest possible root yields a plausible analysis, the algorithm goes with that. Otherwise, it refers to a set of predetermined set of rules (see (Yıldız et al., 2019) for a detailed discussion). In order to use the morphological analysis feature of FrameForm in a different language, Starlang's Turkish Morphological Analysis library can be replaced with a different library pertaining to the target language.

Next step involves the semantic annotation. In this step, the annotator should select the correct meaning of the Lexical Unit in regard to the frame, and appropriate meanings of the other elements in the sample sentence as well (see Figure 2). Au-

⁵<https://github.com/StarlangSoftware/TurkishMorphologicalAnalysis>

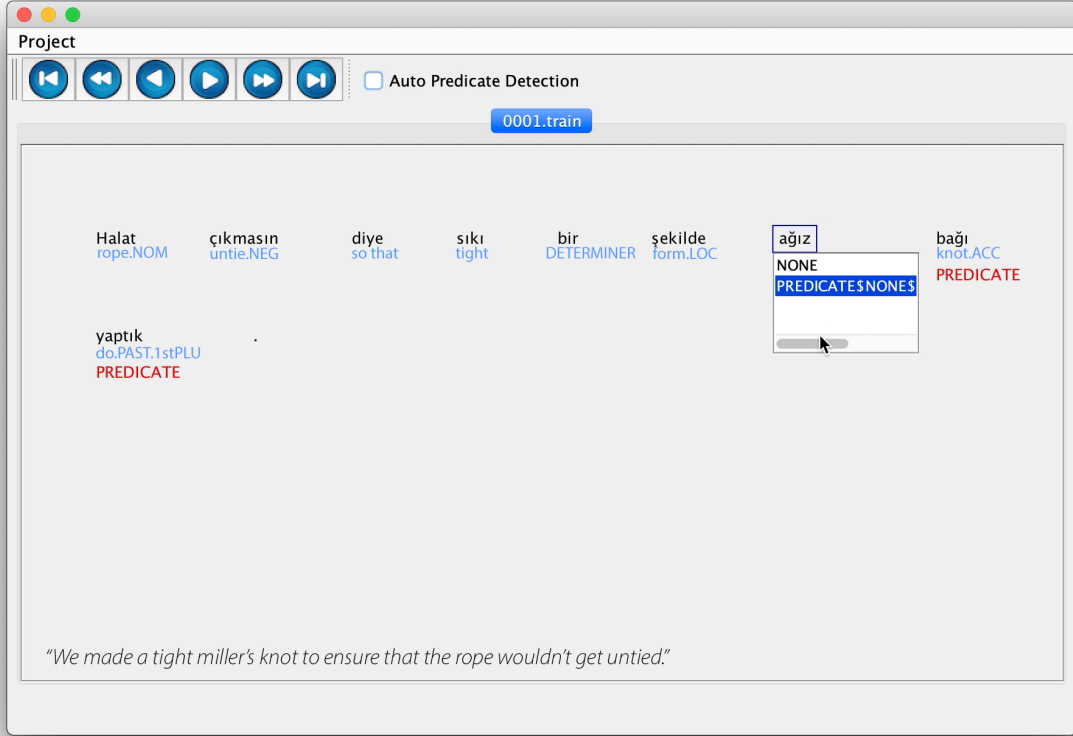


Figure 3: Predicate annotation interface of FrameForm

tomatic semantic annotation is possible in order to make annotation process more seamless but the annotator can always change or manually select the meanings. For the certain multi-word expressions (such as phrasal verbs, idioms, etc.) or the words that have only one meaning, the annotation is done automatically. The rest of the words are annotated by human annotators.

For the semantic annotation step, FrameForm refers to a dictionary or WordNet. For the purposes of Turkish FrameNet, we used Turkish WordNet KeNet (Bakay et al., 2021) in order to make Turkish FrameNet compatible with other resources in Turkish (such as Turkish PropBank (Kara et al., 2020)) yet it is possible to introduce different dictionaries or WordNets in order to use FrameForm in different languages.

After the semantic annotation, the annotator should move on to predicate selection screen where they need to mark the predicate/Lexical Unit in the sample sentence (see Figure 3).

Final step is annotating the Frame Elements where the annotator can see all the FEs within that frame and match them with related sentence ele-

ments (see Figure 4).

One of the most important features of the FrameForm is the fact that it significantly facilitates to ensure inter-annotator agreement and coherency. FrameForm allows all annotators to see each others' annotations, that is why the annotators can discuss specific cases or annotations and notify one another regarding potential agreement issues. In addition, FrameForm groups together the Lexical Units and Frame Elements of a single Frame. This way, the annotators can only see and select the Frame Elements pertaining to the Frame they are annotating. Thus, the annotators cannot use or mark down Frame Elements of other frames.

3.1 Interfaces

FrameForm has 4 different screens for the each step of the annotation process: Morphologic analysis screen (see Figure 1), semantic annotation screen (see Figure 2), predicate marking screen (see Figure 3) and frame element annotation screen (see Figure 4).

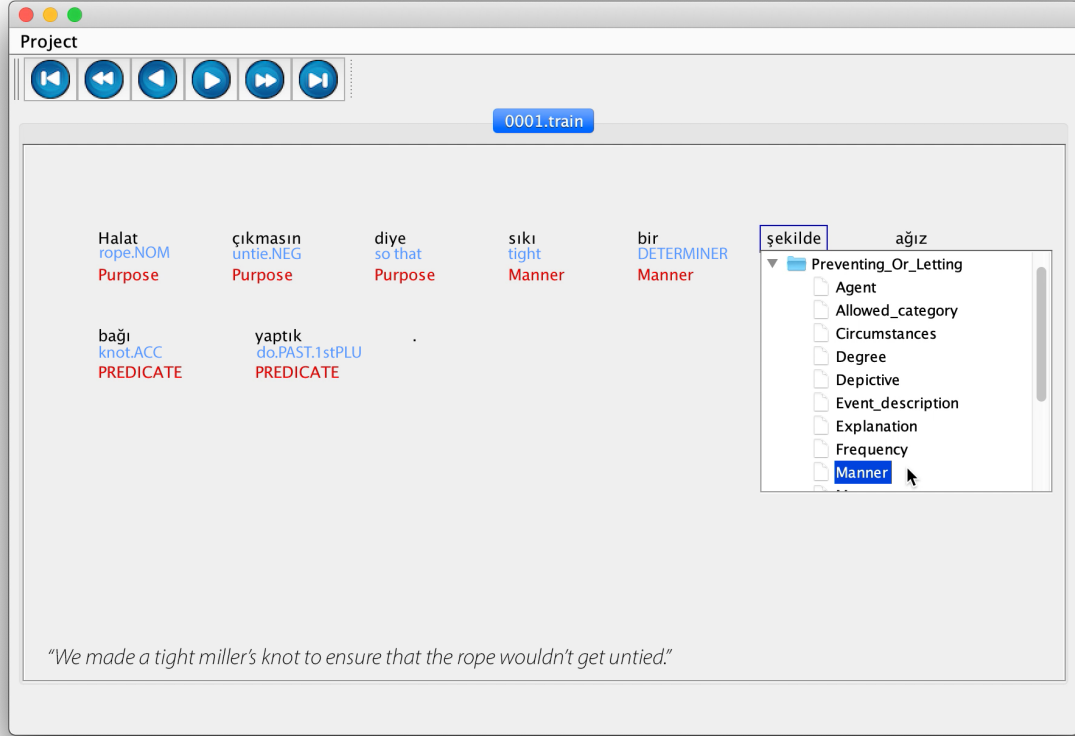


Figure 4: Frame element annotation interface of FrameForm

3.2 What can be annotated with FrameForm?

FrameForm is a very potent tool for annotation. It allows the user to:

- Create new frames,
- Transfer data between the frames,
- Manually edit or change sample sentences,
- Delete Lexical Units,
- Do morphologic analysis, semantic annotation, predicate marking and frame element annotation.

4 Conclusion and Future Studies

With FrameForm, we aimed to create a potent, flexible, easy-to-use annotation tool. In order to ensure that FrameForm alone is enough for every step of the FrameNet annotation and maintenance processes, we equipped our tool with a wide range of features including semantic annotation and Frame Element annotation. Thus it is possible to create

a FrameNet from scratch, grow it and maintain it using only FrameForm.

FrameForm can be downloaded freely on GitHub. Being easy to access and distribute, a crowded team of annotators can use FrameForm for their annotation needs. Since annotators can see progress made by other members of the team, FrameForm makes it easier to ensure inter-annotator agreement.

One of our main goals was to make FrameForm capable of answering the needs of other FrameNet teams. That is why it is an open-source tool that can be modified or advanced in accordance with the unique needs and typologies of other languages. Further studies can focus on the compatibility of FrameForm with other languages and what should be improved.

References

Özge Bakay, Özlem Ergelen, Elif Sarmış, Selin Yıldırım, Bilge Nas Arıcan, Atilla Kocabalcıoğlu, Merve Özçelik, Ezgi Sanıyar, Oğuzhan Kuyrukçu, Begüm Avar, and Olcay Taner Yıldız. 2021. *Turkish WordNet KeNet*. In *Proceedings of the 11th*

- Global Wordnet Conference*, pages 166–174, University of South Africa (UNISA). Global Wordnet Association.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. [The Berkeley FrameNet project](#). In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.
- Lars Borin, Markus Forsberg, Leif-Jöran Olsson, Olof Olsson, and Jonatan Uppström. 2013. The lexical editing system of karp. In *Proceedings of the eLex 2013 conference*, pages 503–516.
- Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, and Sebastian Pado. 2006. [SALTO - a versatile multi-level annotation tool](#). In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).
- Charles J Fillmore and Beryl TS Atkins. 1998. Framenet and lexicographic reference. In *First International Conference on language resources & evaluation: Granada, Spain, 28-30 May 1998*, pages 417–426. European Language Resources Association.
- Charles J. Fillmore, Collin F. Baker, and Hiroaki Sato. 2002. The FrameNet database and software tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).
- Charles J Fillmore et al. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32. New York.
- Christopher Johnson, Charles J Fillmore, E Wood, Josef Ruppenhofer, Margaret Urban, MIRIAM Petruck, and COLLIN Baker. 2001. The framenet project: Tools for lexicon building. *Manuscript. Berkeley, CA, International Computer Science Institute*.
- Neslihan Kara, Deniz Baran Aslan, Büşra Marşan, Özge Bakay, Koray Ak, and Olcay Taner Yıldız. 2020. [TRopBank: Turkish PropBank v2.0](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2763–2772, Marseille, France. European Language Resources Association.
- John B Lowe. 1997. A frame-semantic approach to semantic annotation. In *Tagging Text with Lexical Semantics: Why, What, and How?*
- Büşra Marşan, Neslihan Kara, Merve Özçelik, Bilge Nas Arıcan, Neslihan Cesur, Aslı Kuzgun, Ezgi Sanyar, Oğuzhan Kuyrukçu, and Olcay Taner Yıldız. 2021. [Building the Turkish FrameNet](#). In *Proceedings of the 11th Global Wordnet Conference*, pages 118–125, University of South Africa (UNISA). Global Wordnet Association.
- Ely Matos and Tiago Torrent. 2016. A flexible tool for an enriched framenet. In *Proceedings of the 9th International Conference on Construction Grammar (ICCG9)*, Juiz de Fora, Brazil, october. Federal University of Juiz de Fora (UFJF).
- Josef Ruppenhofer, Jonas Sunde, and Manfred Pinkal. 2010. [Generating FrameNets of various granularities: The FrameNet transformer](#). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Olcay Taner Yıldız, 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)*, pages 1364–1372, Varna, Bulgaria. INCOMA Ltd.