Introduction to Semantic Role Labelling

Behrang QasemiZadeh

zadeh@phil.hhu.de

Computational Linguistics Department, HHU – DRAFT

October 2018–January 2019

Language Resources for Semantic Role Labeling

The next few sessions introduce language resources available for Semantic Role Labeling. Namely, we will look into:

FrameNet

PropBank

VerbNet

(And maybe a few more resources.)

Language Resources for Semantic Role Labeling (contd.)

Language resource

Put simply, a language resource is a *machine readable* source of *linguistic information* which can be used for automatic analyses of natural language or investigating linguistic phenomenon in a systematic manner.

At the core of a language resource is its annotation layers, i.e., systematic labeling of linguistic information which cannot be accessed/inferred directly from the surface structure of sentences.

Simply put, annotations are meta-data for linguistic structures (e.g., part-of-speech tags for words, syntactic analysis for sentences, semantic relation between words, etc.).

Language Resources for Semantic Role Labeling (contd.)

Language resources come in different format and annotation layers (syntax, semantics, discourse, and so on).

For instance, machine-readable dictionaries such as *WordNet* (classifications which are not anchored to contextual usages) and annotated corpora (in which annotations are asserted in context, e.g., part-of-speech tag for words in *sentences*) are the two major classes of language resources.

Some resources stand in between or they provide both a dictionary-like information source and an annotated corpus (e.g., FrameNet).

FrameNet: Before we begin

Before we begin, FrameNet is available online and it can be browsed using frame index:

```
http://corpora.phil.hhu.de/framenet/fndata-1.7/
frameIndex.xml
```

or, using its lexical unit index:

```
http:
//corpora.phil.hhu.de/framenet/fndata-1.7/luIndex.xml
```

Let's have a look at them first and then see what is what!

FrameNet: Intro

As discussed previously, FrameNet has been originated from Fillmore's Frame Semantics:

- Situations/Events are the primary units of meanings;
- Situations involve a number of participants;
- The semantic roles of participants are defined with respect to their semantic-relation to situations.
- Frames (data structure from AI) is used to organize them.

Also, frame is, simply, a data structure consists of a number of frame elements that are packaged and labeled using a controlled vocabulary.

But what does it mean? Let's continue by an example.

FrameNet: Intro (contd.)

Activity_start is a FrameNet frame (we call **Activity_start** a frame type from now on), which is defined as

> *An* *Agent* *initiates the beginning of an ongoing* *Activity* *in which he will be continuously involved.*

In this definition (and all other definitions in FrameNet's website), the colored items identify the Frame Elements (from now on FEs).

The frame definition also include a short description of its FEs.

FrameNet: Intro (contd.)

For example, for the **Activity_start** frame, the FEs are:

CORE: *   Activity[Act]   This FE identifies the Activity that an Agent starts.

          *   Agent[Age]   This FE identifies the being that starts the intentional Activity.

Non-Core: *   Circumstances [Cir]   Circumstances describe the state of the world (at a particular time and place) which is specifically independent of the event itself and any of its participants.

          . . .

FrameNet: Intro (contd.)

Types of FEs

- **Core**: conceptually necessary for the Frame, <u>roughly</u> similar to syntactically obligatory;
- **Non-Core**: FEs that are not central to the frame's intended meaning, but:
    - **Peripheral**: Providing peripheral information such as (sometimes) time and place, manner, etc. (depending on the frame definition ; (<u>approx.</u> like adjuncts);
    - **Extra-thematic**: they are not specific to a frame and relate Frames to the broader context.

FrameNet: Intro (contd.)

We limit our scope to Core FEs. Do we have any **Activity_start** frame in the following sentences (and if yes, what are the Core FEs):

(1)   I started to cry to get out of it.

FrameNet: Intro (contd.)

I [Agent] STARTED to cry [Activity] to get out of it.

which we can, e.g., use the following notation to represent it:

$$
\begin{bmatrix}
\textbf{Activity\_start}.start & \\
\text{AGENT} & I \\
\text{ACTIVITY} & cry
\end{bmatrix}
$$

What about the following sentences, can you identify **Activity_start**?

(2)   The window started to break.

(3)   Price of oil started to increase.

More on FrameNet's Frame

In FrameNet, some of the FEs are accompanied by an additional information under semantic type, e.g., the Agent for the **Activity_start** frame is marked as Semantic Type: *Sentient*.

These are basic semantic typing for FE fillers and have been used for several purposes such as frame parsing (see Ruppenhofer et al. (2016) pp. 79).

* FE fillers: a span of text that "fills" a FE.

More on FrameNet's Frame (contd.)

Lexical unit is another term specific to FrameNet.

lexical unit (LU): a pairing of a lemma and frame: A word form w (of curtain meaning – context is important) brings to mind some frame f. LU is this pairing of w and f. For instance start+**Activity_start**. (**Notice:** word+sense $\neq$ word+frame.)

As far as I understood, neither f or w can be taken as the starting point in FrameNet, they go hand in hand!

There is no guarantee that FrameNet-based word-sense categorization be the same as dictionaries (they are not! differences are interesting).

More on FrameNet's Frame (contd.)

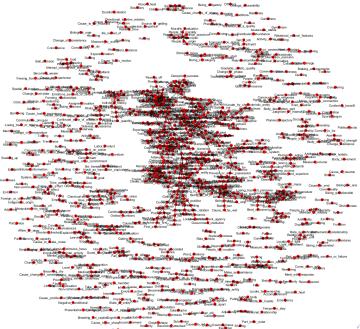**Frame Relationship**

FrameNet defines a network of semantic relationships between frames, too.

The most important relationship is inheritance, in which one frame borrows/inherits definitions/structures from another frame.

For example, **Revenge** inherits from **Rewards_and_Punishment**, **Rewards_and_Punishment** inherits from **Intentionally_affect**, and **Response** from **Eventive_affecting**, which itself inherits from **Intentionally_act**, **Transitive_action** and so on.

The inheritance graph of FrameNet can be particularly an interesting area to analyze and extend: It is complex and incomplete and that makes it even more interesting.

Other Frame-to-frame relations

Apart from inheritance, FrameNet uses a number of other frame-to-frame relations (see Chapter 6 of Ruppenhofer et al. (2016)):

Perspective_on relation is used when one 'Neutral' frame (a unqiue situation) can be interpreted differently wrt. to the points of view on the situation/frame. Tangible example is **Commerce_buy** and **Commerce_sell**, which are both Perspective_on **Commerce_goods-transfer**.

The usages of Perspective_on relation go beyond the our simple example, see Perspective_on the "State of Affairs" frames.

Other Frame-to-frame relations (contd.)

The remaining relations are:

Uses: similar to Perspective_on, see **Judgment_communication** as an example. Mostly used between child and parent frames.

Subframe of: Complex frames can be divided into a sequences of states and transitions. Each of these states can be modelled as a frame. These separate frames (called subframes) are related to the main complex frame using the SubFrame relation. I

Precedes: is defined only between two subframes of their complex frame (look for frames with their name starts with 'Pre_'/'Post_' they are often in a Precedes relation).

Other Frame-to-frame relations (contd.)

Inchoative of and **Causative of**: As their names suggest, one is the inchoative/causative version of the other. Causatives should inherit from the **Transitive_action**, Inchoatives should inherit from **Event**, and States from **State** or **Gradable_attributes**.

(examples to follow.)

More examples

We introduced the frame **Activity_start**, which inherits from **Process_start**.

Process_start definition:

> An `Event` begins at a certain `Time` and `Place`. Reason may also be indicated. NB: Refer to Event frame.

`Event` is the core FE of **Process_start** :

> Name of the Event which occurs. (Semantic Type: State_of_affairs)

More examples (contd.)

Cause_to_start is another frame which is in *Causative of* relation to **Process_start**. The definition of **Cause_to_start**:

> A `Cause`, animate or inanimate, causes a process, the `Effect`, to begin.

in which

`Cause`: An animate or inanimate entity, a force, or event that produces an effect. Volitionality is not a necessary characteristic of Causes.

`Effect`: This Frame Element marks expressions that indicate the process that the Cause initiates.

More examples (contd.)

What is your analysis (which frame) for:

(4)   That was when all the fun started.

(5)   The window started to break.

(6)   Price of oil started to increase.

(7)   Lightning started a number of small forest fires.

(8)   John started a new company.

(9)   John started to dance.

Additional Info – Non-Lexical Frame

Note that several types of frames exist in FrameNet:

```
Human, Visual_modality, Transparent Noun, Activity,
Artifact, Negative, Non-perspectivalized_frame, State,
Non-Lexical Frame, Event, Locative_relation,
End_of_scale, Physical_object
```

Particularly, *Non-Lexical Frame* are abstract frames which are not evoked and lexicalized by words directly, but through their related frames.

Additional Info – Non-Lexical Frame (contd.)

Some statistics (v 1.7)

FrameNet contains 1,221 entries for frame definitions.

In the definitions of these frames, FrameNet introduces 1,285 frame elements (semantic roles) of which 1002 are CORE elements (notice that most element types are CORE).

It covers 13,631 LUs for 10,481 lemmas. Top 10 frames with largest LU set:

Frame	#LUs	Frame	#LUs
Emotion_directed	186	Containers	120
Stimulus_focus	179	Natural_features	115
Clothing	163	Leadership	106
Self_motion	152	Make_noise	105
Experiencer_obj	136	Buildings	104

A Summary So Far: Lexicographic Perspective

Put simply, FrameNet can be seen as a dictionary of frame-situations.

To understand meanings of words, we use a system proposed in most dictionaries:

- A dictionary has one **entry** per word/lexical form.
- Entries are sorted in an alphabetical order.
- In each entry for the word, the meanings of words (definition for word senses) are listed, usually followed by an example.
- Additionally, lexical semantic relationships (synonymy, antonymy, etc.) connect and organize word-meaning pairs.

A Summary So Far: Lexicographic Perspective (contd.)

In FrameNet, the situation is analogous (not the same):

- We have a number of situation Frames. These are defined and added with respect to some guidelines (see Ruppenhofer et al. (2016)).
    - The definition of frames consists of a description of the situation, and its participants.
- Words are paired with **FRAME**s that they can <u>evoke</u> in a context (the Lexical Units).
- Frames are related to each other through a number of frame-to-frame relationships, as mentioned briefly.

A Summary So Far: Lexicographic Perspective (contd.)

Topics for you to think and hopefully undertake to tackle as a project work:

- What are the best methods and practices for finding the frame that is evoked by a word in a context?

    + We have word-frame inverted indexes but it is not enough in many situations.
    + There is much more in FrameNet than LUs (e.g., frame-to-frame relations), how to use them for finding frames?
    + How to use a machine learning algorithms in an interactive way? Linguistic-oriented method, etc.? Ideas from prototype?!

- How to make sure a chosen frame for a word is the correct one?

- FrameNet is not complete and it may never be (by its nature). How to make sure that FrameNet does not contain a suitable Frame for a word given in a context?

FrameNet Full Text Annotations

Apart from definitions, relations, LUs, etc., FrameNet contains full text annotations.

Simply put, some of lexical items (mostly verbs) in several sentences are annotated (the source of sentences are stated in the FrameNet fulltext annotation package – mostly ANC, WSJ, and more news stuff, which are added to the original BNC examples).

"full text annotations" does not mean that all words in the sentence are annotated (i.e., it is not frame-based analysis of sentences).

The full text annotations can be browsed online, or you can register and obtain a copy for yourself, which will let you process and manipulate FrameNet data in XML files.

This data has been used for developing many Frame parsers.

FrameNet in Other Languages

FrameNet is available in (or work in-progress) several languages other than English. For German, I am not sure (if you find a working resource please let us know).

Bibliography

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., and Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. ICSI, Berkeley.