Verb alternations and their impact on frame induction

Esther Seyffarth

Heinrich Heine University, Düsseldorf

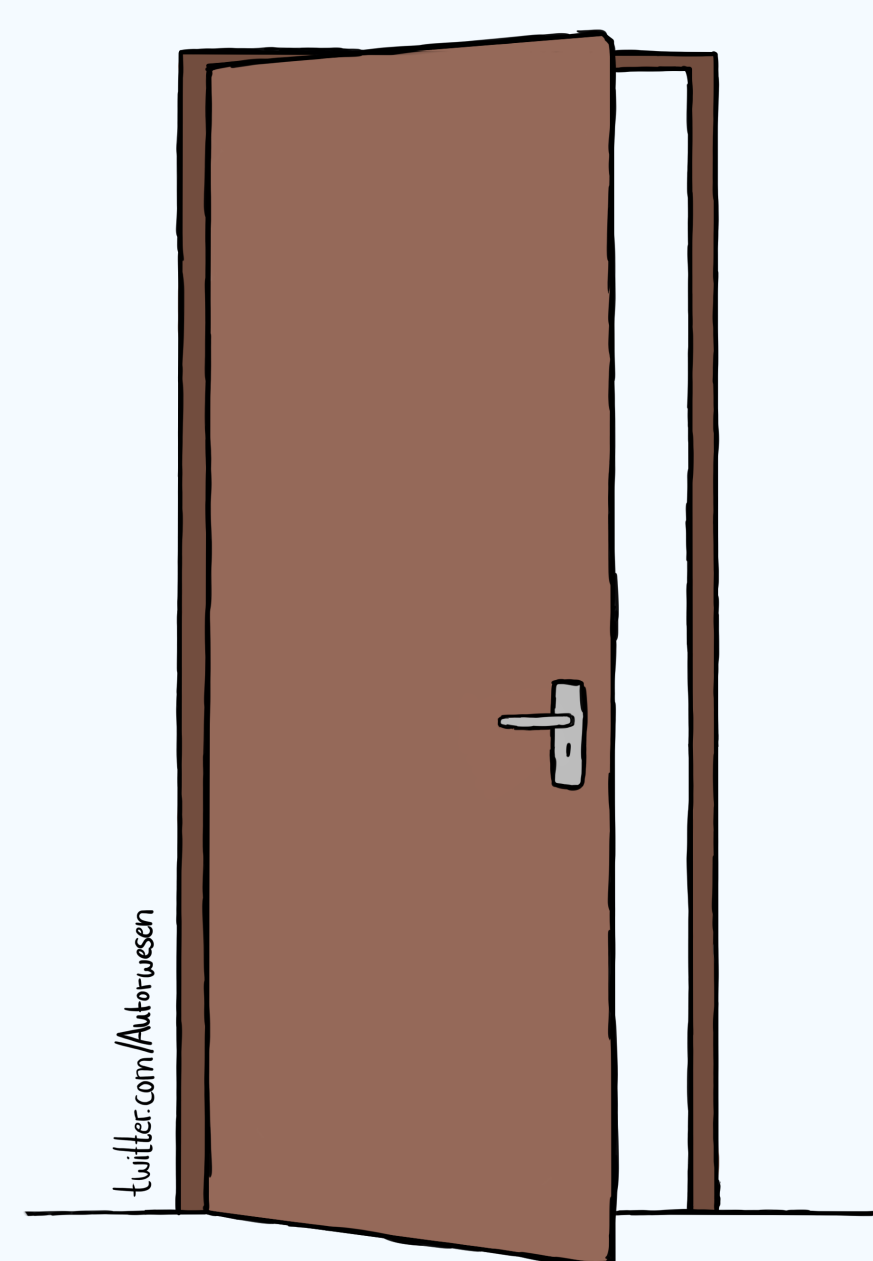HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Verb Alternations

- Verbs that participate in **diathesis alternations** (Levin, 1993) can occur with more than one valency pattern; in many cases, switching between different syntactic realizations systematically changes the semantics of the verb.
- Here, we are interested in alternating verbs that denote events, where different uses of the verbs denote different types of events.
- With different sets of syntactic arguments, alternating verbs can specify different sets of semantic roles involved in the event they denote.
- Examples for English verb alternations are **causative** alternations, the conative alternation, or the dative alternation.
- Verb alternations are interesting from a computational linguistics point of view: Semantic representations, and NLP tools in general, must take differences and similarities between alternating forms into account.
- Alternating verbs are **not homographs**: Their meanings are systematically related to each other.
- Alternating verbs are **not synonyms**: Their meanings are systematically different from each other.
- Example for the causative alternation: **intransitive** *open* and **transitive** *open* (see below).

Frame Induction

- **Semantic frames** (Fillmore, 1968; Minsky, 1974; Barsalou, 1992) are recursive attribute-value structures that represent concepts.
- Concepts can be described in terms of the attributes that contribute to their meaning. Event frames can specify the semantic roles that are involved in the event.
- (Functional) relationships between concepts and their attributes represent the relationships that exist between entities and their properties.
- Frame-semantic lexical resources like FrameNet (Ruppenhofer et al., 2006) or PropBank (Palmer et al., 2005) map lexical units of a language to frames.
- There is a many-to-many relationship between lexical units and semantic frames.
- **Frame induction** is the process of deriving frame structures and hierarchies from text:
  - Which frames can a verb evoke?
  - Which verbs evoke the same frame?
  - Which arguments are part of a particular frame?
  - Which argument slots (role slots) are obligatory, and which ones are optional?
- Creating frame resources automatically (unsupervised or semi-supervised) can be more sustainable than creating them manually. A frame induction system should be transferable and adaptable to other data sources and/or languages.

Proposed frames for alternating verbs

(1) The door opens.

(2) Alex opens the door.

$$\begin{bmatrix} open \\ \text{THEME} & \text{the door} \end{bmatrix}$$

$$\begin{bmatrix} causation \\ \text{CAUSE} & \begin{bmatrix} activity \\ \text{AGENT} & \text{Alex} \end{bmatrix} \\ \text{EFFECT} & \begin{bmatrix} open \\ \text{THEME} & \text{the door} \end{bmatrix} \end{bmatrix}$$

Discussion

- The frames proposed here encode the alternating meanings in the form of embedded structures: The added causative meaning of the transitive sentence in (2) "wraps around" the meaning of the intransitive sentence in (1).
- With this analysis, the semantic representation of (2) is identical to an analysis of the paraphrase *Alex causes the door to open*.
- This structure makes it possible to generate frame representations systematically, if a verb is known to participate in this type of alternation.
- The **identification of alternating verbs** is one of the tasks the frame inducer needs to fulfill in order to generate these frames.
- Some other alternations can also be modelled in a similar fashion. For instance, a frame for the conative alternation may look like the following, for the sentence *Kim cut at the rope*:

$$\begin{bmatrix} attempt \\ \text{AGENT} & \text{Kim} \\ \text{ACTION} & \begin{bmatrix} cut \\ \text{AGENT} & \text{Kim} \\ \text{THEME} & \text{the rope} \end{bmatrix} \end{bmatrix}$$

- Not all alternations are good candidates for embedding frames. For instance, in the dative alternation, different uses of a verb evoke different frames; it is not possible to systematically embed one of the frames in the other, as in the causative alternation:

  (3) Kim gives the book to Pat. (frame: *change-of-location*)

  (4) Kim gives Pat the book. (frame: *change-of-possession*)

- Without a linguistic account of verb alternations, a frame inducer can only classify alternating verb frames as identical (like synonyms) or distinct (like homographs). Both options lead to an incomplete picture. **This is the main motivation for our ongoing investigation of these phenomena.**
- One of the goals of frame induction is the creation of frame resources that can be used in similar ways as the handcrafted resources mentioned above. In addition to this, the resulting resource also provides additional empirical evidence for the existence and the nature of frames.

Related work and future plans

- The frames in FrameNet/PropBank for alternating verbs differ from ours. FrameNet does not embed one frame in the other; PropBank treats alternating verbs like synonyms.
- The frame structures presented above are inspired by the work of Osswald and Van Valin Jr (2014), who focus more on a decompositional structure. Their fully decompositional approach is useful from a linguistic perspective, but difficult to implement in a computational setting; the more detailed the semantic representation is, the harder it is to learn automatically.
- Our lab's ***SEM** paper on frame induction: *Coarse Lexical Frame Acquisition at the Syntax-Semantics Interface Using a Latent-Variable PCFG Model* (Kallmeyer et al., 2018; presented on June 5th in Session 4)
- In my PhD thesis, I am working on verb alternations, exploring options for the automatic identification of alternating verbs, and developing strategies for the semantic modelling of alternations.

References (selection)

Barsalou, L. (1992). Frames, concepts, and conceptual fields.

Fillmore, C. J. (1968). The Case for Case.

Kallmeyer, L., QasemiZadeh, B., Cheung, J. C. K. (2018). Coarse Lexical Frame Acquisition at the Syntax-Semantics Interface Using a Latent-Variable PCFG Model.

Levin, B. (1993). English Verb Classes and Alternations: A Preliminary Investigation.

McCarthy, D. (2001). Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences.

McCarthy, D., Korhonen, A. (1998). Detecting Verbal Participation in Diathesis Alternations.

Minsky, M. (1974). A Framework for Representing Knowledge.

Osswald, R., van Valin Jr., R. (2014). FrameNet, Frame Structure, and the Syntax-Semantics Interface.

Palmer, M., Gildea, D., Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles.

Rappaport Hovav, M. (2014). Lexical content and context: The causative alternation in English revisited.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Scheffczyk, J. (2006). FrameNet II: Extended Theory and Practice.

See proceedings for full bibliography.

NAACL HLT