Advances in Cognitive Systems Student Research Workshop (2015) 1–4 Submitted 4/2015; published X/20XX

An Analogical Approach to Learning Constraints on Non-canonical Constructions

Clifton McFate C-MCFATE@NORTHWESTERN.EDU
QRG Lab, Northwestern University, Evanston, IL. 60201 USA

Abstract

Non-canonical constructions differ from canonical structures in that they carry extra pragmatic meaning relating their arguments to the current discourse. One important question is how language users learn the pragmatic relationships that govern these variations. This paper proposes that these relationships can be learned through analogical generalization over first explicit and then inferentially related examples. We focus on the preposing construction and simulate this learning process for a proof of concept example using the SAGE model of analogical generalization.

1. Introduction

English offers a myriad of ways to express identical propositional content. For example, the following sentences express the same simple relationship between a pot and a table: a) A pot sat on the table. b) On the table there sat a pot. c) It was a pot that sat on the table.
Birner & Ward's (1998) analysis demonstrates that non-canonical variations carry restrictions on information status that the canonical construction does not.

While it's clear that non-canonical word order (NWO) constructions follow a different distributional pattern than their canonical (CO) counterparts, what's less clear is how humans are able to learn these distinctions. One possibility could be generalization through structural alignment. Gentner's (1983) structure mapping process provides a powerful mechanism for abstracting rule-like structures from individual examples (Gentner & Medina, 1998). This paper suggests the same mechanism could account for learning the relationships which constrain these constructions. This claim is evaluated with a proof of concept example using the SAGE model.

2. Background

2.1 Preposing and Non-canonical Word Order

In NWO preposing constructions, a normally post-verbal argument in CO instead appears pre-verbally. Consider the preposing examples in 1(b, d), all of which are felicitous following the question "Are you as addicted to Game of Thrones as I am?". 1(a, c) are canonical word order.

1) a) You're addicted to Game of Thrones?
   b) Game of Thrones you're addicted to?
   c) Yeah. I find Game of Thrones thrilling.
   d) Yeah. Game of Thrones I find thrilling!

Birner and Ward (1998) characterize the constraints on preposed arguments by expanding on Prince's (1992) definition of discourse-old. Per Prince, an entity is discourse-old if it was previously mentioned in the discourse and discourse-new otherwise. Birner and Ward (1998) extend the definition of discourse-old to include entities that are 'inferrable' based on the existing discourse. They argued that relations which allow preposing are relations of partially ordered sets (poset) such as part/whole, type/subtype, or identity. The examples above all involved identity. Consider the examples in 2 as responses to the same question.

2. a) I love Game of Thrones. So much gore it has!
   b) Kind of. Boardwalk Empire I much prefer.

In 2a, the preposed constituent is an attribute of the triggering set. In 2b, the relation is type/subtype. Constituents that aren't related to the previous discourse by a salient poset relationship can't be preposed.

3. a) I just bought the DVD. # My doctor I have to call.
   b) I love Game of Thrones. # Such a cold climate it has.

In 3a, the doctor is not part of a standard poset relating to purchasing a movie. Similarly, in 3b climate is not a standard attribute of a television show. Thus, preposed constituents must be discourse-old. However, a complex set of inferrable relationships can grant discourse-old status. How then could humans and machines learn this set of inferrable relationships? One method is analogical generalization over naturally occurring examples.

2.2 Analogical Generalization

Gentner and Medina (1998) propose that structural alignment facilitates learning by highlighting shared systems of relations for generalization across examples. Comparison is certainly useful for learning linguistic mappings. Christie and Gentner (2010) demonstrated that comparing two examples allowed children to extend a relational novel label.

Structure mapping has been implemented in a computational model, Falkenhainer *et al*'s (1989) structure mapping engine (SME). SME takes a base and target case of structured representations and produces mappings. Relationships that are present in the base but missing in the target can be hypothesized as candidate inferences.

The Sequential Analogical Generalization Engine (SAGE) uses SME and analogical retrieval to generalize from examples (McClure et al, 2015). SAGE operates over structured representations. First a relevant generalization (if any) is retrieved, and then the retrieved case is generalized to incorporate the new example. Where expressions match, SAGE updates their probabilities. When matching expressions have non-identical entities, SAGE creates a generalized entity with a probability distribution representing possible instantiations.

3. Analogical Generalization of Preposed Constructions

Analogical generalization, operating over many individual examples of preposing, could result in a schema-like structure which represents their discourse-constraints as a set of likely inter-sentential relationships. A good starting place for generalization would be preposed constituents licensed by lexical identity. Identity and similar syntax would provide the shared structure needed to begin generalization. Over the course of several similar examples, the generalization abstracts away from the specific entities while retaining the crucial co-reference. When a new example occurs with a different linking relation, it gets added to the generalization but does not have the identity relationship. At this point, either an evoked relationship is added instead, or a new relationship is hypothesized via candidate inference.

3.1 Proof of Concept

SAGE was used to construct a generalization over examples of simple preposing constructions. Each was of the form PP-NP-VP (e.g. 'In the yard a cat sat') and followed a transitive-pp sentence where the prepositional phrase contained the same lexical item as the preposed constituent. These sentences were represented in Cyc[1] predicate calculus. The syntax of each sentence was represented using de Marneffe *et al* (2006) dependency parse representations from the Stanford parser. This representation was augmented to include `before` predicates which hold between NP and VP heads and captures the word order differences. The identically relationship was represented with a `coref` predicate.

Each pairing of an initial SVO sentence and a preposed PP-NP-VP sentence was represented as a single case. These cases are then sequentially fed into SAGE for generalization. After three training cases, SAGE successfully creates a single generalization. A portion of this generalization is shown in *figure 1***.**

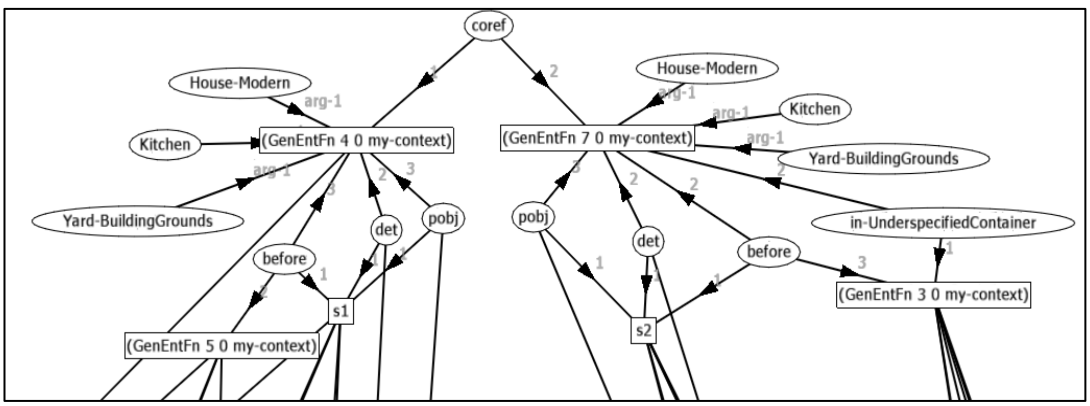

Figure 1: Partial Graph of SAGE generalization

In figure 1, the focused relationship is the licensing `coref` lexical identity predicate. It holds between two generalized entities which have various traits. Definitional information about the entities (e.g. house, yard, and kitchen) are all alternatives to the same relation.

Now, a fourth example with a part/whole relationship is added. The shared syntactic structure allows this example to be grouped into the existing generalization, and over time, the relevant preposing constraints will emerge as likely relationships between the preposed entity and a reference in the prior discourse.

(partWhole (GenEntFn 4 0 my-context) (GenEntFn 7 0 my-context))	0.25
(coref (GenEntFn 4 0 my-context) (GenEntFn 7 0 my-context))	0.75

[1] http://www.cyc.com/kb

If we don't assume that the relevant relationship is represented, the task becomes more difficult. Instead, the system could learn the relationship as a candidate inference. When a case missing the `partWhole` relationship is compared via SME to the generalization, it generates the following candidate inference (among others): `(coref kitchen100 fridge200)`. This approach would require a consistency check to reject the lexical identity assertion and look for an alternative. This could also be represented by using a placeholder function for `coref` and allowing the system to accumulate evidence as to what type of relationship it is over time.

4. Conclusions

I propose that analogical generalization can be used to learn pragmatic constraints on preposing constructions. Analogical generalization provides a powerful domain independent learning mechanism that can learn abstract concepts with limited input and little feedback. This makes it ideal for learning linguistic generalizations. One possible path for this process involves analogical generalization over constructions with similar syntax and a strong lexical identity relationship. Over time, this generalization would provide the scaffolding needed to learn the more abstract linking relationships that govern preposing.

Much work remains to be done. For example, it is unclear whether or not the poset relationships are evoked as a part of world knowledge or discovered as a result of candidate inference. Furthermore, it's possible that generalizations become a part of semantic parsing, or it's possible they are used after parsing to check for coherence. However, ultimately, analogical generalization provides a strong mechanism for learning representations that address pragmatic, syntactic, and semantic constraints.

References

Birner, B. J., & Ward, G. (1998). *Information Status and Noncanonical Word Order in English* (Vol. 40). John Benjamins Publishing.

de Marneffe, M., MacCartney, B., & Manning, D. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In *LREC* 2006.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence, 41*(1), 1-63.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive science*, 7(2), 155-170.

Gentner, D. & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, 2(2). 261-283.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65(2)

McLure, M.D., Friedman S.E. and Forbus, K.D. (2015). Extending Analogical Generalization with Near-Misses. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, Texas

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In Sandra Thompson and William Mann (eds.), *Discourse Description: Diverse Analyses of a Fundraising Text*, 295-325. Amsterdam, John Benjamins.