

Jackendoffian Tiers and Atomese Semantics

Ben Goertzel

February 12, 2013

The semantic interpretation of any given natural language sentence can be expressed in a variety of different ways. In the OpenCog framework it can most naturally be envisioned as a list of relationships, each having one or more arguments – or in other words, a subgraph of the Atomspace. However, for purposes of efficient pattern mining, reasoning, query processing and language generation, it may be valuable to structure semantic interpretations in a more regimented way than is implied by saying “semantic interpretations are subgraphs of the Atomspace.” The challenge is to do this without imposing such a regimented structure as to prevent the expression of the full scope of natural language semantics, or to render overly length or complex the expression of semantic structures commonplace in natural language.

One useful way of organizing Atom structures representing natural language semantics, may be to introduce Jackendoff’s notion of “tiers.” In Jackendoff’s theory of Conceptual Semantics, the semantics of a sentence is decomposed into multiple tiers, e.g.

- The **action tier**, focused on the roles Actor, Undergoer, Patient, and Beneficiary
- The **thematic tier**, focused on the roles Agent, Theme, Location, Goal, Source and Route
- The **referential tier**, that encodes existential claims about the entities involved in a sentence

The way these tiers are manifested in Conceptual Semantics utilizes a lot of highly specific formalism that would be tricky to port to the OpenCog context. However, the basic concept of the three tiers may still be useful in an OpenCog context.

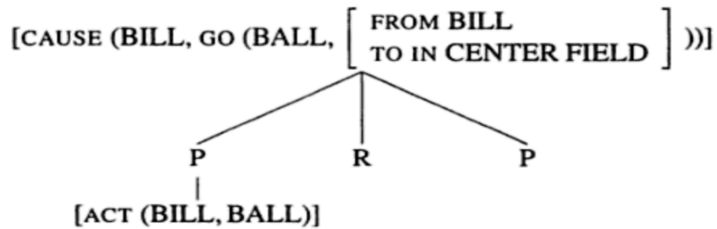
What I will argue here is that it’s possible to use Jackendoff’s tiers to subdivide the OpenCog Atomese interpretation of a sentence into portions, in a way that should be useful for guiding pattern mining and reasoning over natural language interpretations and inferences derived therefrom.

These are rough notes, intended to spur further discussion and research.

Examples of Jackendoffian Tiers

An example of Jackendoff's analysis of a sentence in terms of multiple tiers is given below:

Bill threw the ball into center field.



Here

- the ACT relation is part of the **action** tier
- The CAUSE, GO FROM, TO and IN relations are part of the **thematic** tier (being related to the source, goal and path of the events depicted in the sentence)
- The P link (P for “process”) is part of the **referential** tier, indicating a temporal referential relationship

For another example: In Jackendoff's formulation, the thematic tier of the lexical entry for “drink” is supposed to be as follows:

[Event CAUSE ([Thing]_i, [Event GO ([Thing LIQUID]_j,
[Path TO ([Place IN ([Thing MOUTH OF ([Thing]_i))]])]])])]

As is the tradition in non-computational linguistics, the rules for applying Jackendoff's formalism to natural language sentences are not that precisely defined, and there are plenty of borderline cases for linguistics to argue about in their research papers. However, the broad notion of the three tiers is conceptually inspirational, and it seems feasible to borrow the general concept for OpenCog, while avoiding the particulars of Jackendoff's semantics.

Atomese Analogues of Jackendoffian Tiers

The RelEx interpretation of the above sentence “Bill threw the ball into center field” would look like:

```
_obj(throw, ball)
_subj(throw, Bill)
tense(throw, past)
```

```
into(throw, field)
amod(field, center)
```

and one Atomese interpretation of this would be

```
InheritanceLink Throw_1 Throw
InheritanceLink Ball_1 Ball
InheritanceLink Field_1 Field
InheritanceLink Bill_1 Bill
EvaluationLink Throw_1 (Bill_1 Ball_1)
IntoLink Throw_1 Field_1
InheritanceLink Field_1 Center
TenseLink Throw_1 Past
```

Another Atomese interpretation of this would be

```
(ExistsLink b, B, t, f
  (AndLink
    (InheritanceLink b ball)
    (InheritanceLink B Bill)
    (InheritanceLink t throw)
    (InheritanceLink f field)
    (EvaluationLink t (B, b) )
    (IntoLink t f)
    (InheritanceLink f Center)
    (TenseLink t Past)
  )
)
```

)

Other interpretations such as

```
(ExistsLink b, B, t
  (AndLink
    (InheritanceLink b ball)
    (InheritanceLink B Bill)
    (InheritanceLink t throw)
    (EvaluationLink t (B, b) )
    (IntoLink t Field_1)
    (InheritanceLink Field_1 Center)
    (InheritanceLink Field_1 Field)
    (TenseLink t Past)
  )
)
```

are also perfectly sensible. Atomese contains no requirement that all the terms in an expression be treated the same way. For various linguistic and inferential purposes it may sometimes be most convenient to treat them differently.

The key thing to note about these various interpretations, in terms of Jackendoff's tiers, is that they only differ in certain aspects of the referential tier.

The ACTION tier of the Atom interpretation of the sentence is:

```
EvaluationLink Arg_1 (Arg_2 Arg_3)
```

where in every case we have

```
Inheritance Arg_1 Throw
```

```
Inheritance Arg_2 Bill
```

```
Inheritance Arg_3 Ball
```

The THEMATIC tier of the Atom interpretation of the sentence is, proximally,

Into(Arg_4, Arg_5)

Where in each case we have

Inheritance Arg_4 Throw

Inheritance Arg_5 Field

Inheritance Arg_5 Center

Indirect inference from the Atomese directly generated by the sentence, also generates the THEMATIC relationship

Cause(Arg_2, Into(Arg_4, Arg_5)

i.e.

Cause(Bill, Into(Throw, Field))

-- and this inference would hold true for every interpretation of the sentence, based on reasonable commonsense knowledge about throwing being a causal action.

The REFERENTIAL tier, in all the interpretations, contains the temporal positioning

TenseLink t Past

and apart from that, differs from one interpretation to the other, via differences in quantification.

What this simple example reveals is that Jackendoff's notion of action, thematic and referential tiers corresponds fairly simply to specific Atomese link types. Thus, given an Atom structure produced via natural language understanding, it appears generally feasible to decompose this Atom structure into three Jackendoffian tiers, via apportioning different links to different tiers.

Initial, incomplete rules of thumb for doing this would be:

- EvaluationLinks involving predicates mapped from verbs, usually belong to the Action tier
- Link types derived from prepositions, or EvaluationLinks involving predicates mapped from prepositions, usually belong to the Thematic tier
- Quantifier links, and temporal and spatial links, usually belong to the Referential tier

What is the purpose of this sort of decomposition? It should be useful for guiding operations like pattern mining and reasoning. For some cases of pattern mining and reasoning, we may wish to pay attention solely to the Action tier. For other cases, we may be interested solely in the Thematic tier, or in both the Thematic and Action tiers. For advanced reasoning, we need to use all the tiers. For some sorts of spatial or temporal pattern mining, we need the spatiotemporal portions of the Referential tier, but not the more abstract quantifier portions.