Senseval 3 Logic Forms: A System and Possible Improvements

Altaf Mohammed, Dan Moldovan, and Paul Parker
Language Computer Corporation
Richardson, TX 75080
{altaf, moldovan, parker}@languagecomputer.com

1 Introduction

Logic Forms are first order logic representations of natural language text. The notation is very close to the natural language. A Logic Form is a collection of predicate instances derived from text. A detailed description of the notation is presented in Moldovan and Rus (2001).

Logic Forms can be utilized by a wide variety of applications. A Logic Prover (Rus, 2002; Moldovan et al., 2003) utilizing the axioms generated by the Logic Form generation system boosts the performance of the Question Answering system. The Prover essentially takes as input the Logic Forms of the question and one or more answers and then proceeds to justify (and rank) the answers based on (i) world knowledge axioms, and (ii) NLP axioms. The Logic Prover developed at Language Computer Corporation has increased the performance of the QA system by 30%.

2 Automatic Generation of Logic Forms

2.1 Parse Tree Construction

Logic Forms are derived from the output of a syntactic parser. The first step is the identification of word collocations (based on those identified by WordNet (Miller, 1995)). The parser then proceeds to identify (i) parts of speech of individual words, and (ii) syntactic structure of the text, based on grammar rules. It also differentiates active verb constructs from passive ones. The output is a parse tree. We also include in the parse tree (i) word senses, based on WordNet, and (ii) named-entity tags (Surdeanu and Harabagiu, 2002). Named-entity tags are tags associated with a word or group of words, indicating that they belong to a particular category, for instance, currency, time, date, place, human, etc. The word senses from the parse tree are simply included in the Logic Form as is (for subsequent use in applications), while the named-entity tags are additionally used in the generation of Logic Forms.

2.2 Logic Form Generation

First we identify independent arguments (those arguments which are generated anew for certain predicates). These include arguments for nouns,
verbs (action/eventuality only), and compound
nouns and coordinating conjunctions (for both of
these, only the result argument (or the first one)).
Independent arguments are also generated for
certain adjectival phrases and/or determiners in
the rare cases when they do not qualify a noun
phrase, but stand all by themselves. The inde-
pendent arguments, once generated, are propa-
gated up the parse tree (as predicates form the
leaves of the parse tree). In this way, heads of
phrases are marked.

Next comes the identification of dependent ar-
uments (those which are derived from other in-
dependent and/or dependent arguments). These
may include arguments for modifiers (adjectives,
adverbs), secondary verb slots (all except the
first) and secondary coordinating conjunction
slots (all except the resulting argument), or link-
ing words (prepositions, subordinating conjunc-
tions, etc). The derivation of these arguments
follows from a slot-filling approach and is based
on the interpretation of the parse tree structure
and the associated transformation rule (Moldovan
and Rus, 2001). This is a rule that says how a
particular parse tree structure must be handled,
for instance, 'S -> NP VP' says that the subject of
the main/action verb of VP is the head of phrase
of NP.

3 Dealing with Ambiguous Structures

Named-entity tags are helpful when parsing cer-
tain ambiguous structures. Take, for instance, the
following two sentences:

(i) They gave the visiting team a heavy loss.
(ii) They played football every evening.

The grammar rule for the verb phrase in both
sentences is 'VP -> VB NP NP'.

Whereas, in sentence (i), the first NP is the in-
direct object and the second one the direct, ex-
actly the converse is true for sentence (ii). Upon
closer examination, it is found that 'every eve-
ning' being an indicator of time, does not qualify
for the position of the direct object.

The named-entity recognition system marks all
such indicators of time/date. This enables us to
disqualify these noun phrases as candidates for
the position of the direct object of a verb.

The aforementioned is just one kind of ambi-
guity that we have addressed. Another kind of
ambiguity that we encountered (but have not im-
plemented a solution for yet) is the reduction of
certain words to base forms. Consider, for in-
stance,

(i) John found the key. ('find' in VBN form)
(ii) The King promised to found a similar insti-
tution. ('found' in VB form)

A possible solution we considered was looking
at part-of-speech tags (VBN vs. VB) to resolve
ambiguity, but have not pursued this further in
the (current) absence of a database that maintains
a mapping from inflected word and part-of-
speech pairs to the corresponding base forms.

Another approach considered was choosing the
base form whose frequency of occurrence in the
Brown corpus, as reflected in WordNet, was
highest.

4 Changes/Improvements for Senseval 3

Since no complete specification was given for the
proper formation of logic forms for many special
cases, we chose to model our Senseval 3 Logic
Form system on the provided examples. The LF
system was updated to model the Senseval 3 be-
havior in the following ways.

4.1 Adverbs Modifying Adjectives

These adverbs are assigned the same argument as
the adjective they modify (Mohammed, 2003). For
instance, "the extremely fast athlete" is rep-
resented as "extremely:r_ (x1) fast:a_ (x1) ath-
lete:n_ (x1)".

4.2 Variable Slots for Verbs

The verbs are now given a variable number of
arguments (minimum two: the action/eventuality
and the subject). They get arguments for all verb objects, including prepositional attachments.

Previously, we had a fixed slot allocation mechanism for verbs, specifying always the action, the subject, and the direct object. These slots were filled with dummy arguments in the absence of proper arguments.

Example:

*S: John gave Mary the book on Saturday.*

LF (previous notation)

\[
\text{John}:n_\text{(x1)} \text{give}:v_\text{e1, x1, x3) Mary}:n_\text{(x2) book}:n_\text{(x3) on (e1, x4) Saturday}:n_\text{(x4)}
\]

LF (Senseval 3 notation)

\[
\text{John}:n_\text{(x1)} \text{give}:v_\text{e1, x1, x3, x2, x4) Mary}:n_\text{(x2) book}:n_\text{(x3) on (e1, x4) Saturday}:n_\text{(x4)}
\]

### 4.3 Subordinating Conjunctions

These conjunctions are given two arguments. The second argument is the main/action verb of the subordinate clause. The first argument is assigned as follows: (i) if the clause attaches to a sentence (or a verb phrase), then the main/action verb of this sentence (or verb phrase), (ii) if the clause attaches to a noun phrase, then the head of the noun phrase. Additional details are presented in Mohammed (2003).

Example:

*If you heat ice, it melts.*

LF:

\[
\text{If (e2, e1) you (x1) heat}:v_\text{e1, x1, x2) ice}:n_\text{(x2) it (x3) melt}:v_\text{e2, x3)}
\]

### 5 Impact of Parse Tree Accuracy on Logic Forms

The Logic Forms are derived directly from the parse trees. This makes the generation of accurate parse trees extremely important. We have analyzed the performance of automatically generated Logic Forms based on both the machine-generated (hence necessarily somewhat erroneous) parse trees and parse trees generated by human annotators.

The following results are based on the set of 300 test sentences provided for the Logic Forms Identification task at Senseval 3. The number of sentences with all predicates correctly identified has increased from 155 to 191, an improvement of 23.2%. The number of sentences with all correct arguments and all correct predicates (in other words, 100% correct Logic Forms) has increased from 65 to 98, a 50.7% improvement over Logic Forms derived from machine-generated parse trees. The results are presented in Table 1. Note that the row captioned “Predicates” indicates the number of sentences for which all predicates were correctly identified, while the row captioned “Entire LF” indicates those for which all predicates as well as all associated arguments were correctly identified.

<table>
<thead>
<tr>
<th></th>
<th>Machine Parse</th>
<th>Hand Parse</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicates</strong></td>
<td>155 / 300</td>
<td>191 / 300</td>
<td>23.2%</td>
</tr>
<tr>
<td><strong>Entire LF</strong></td>
<td>65 / 300</td>
<td>98 / 300</td>
<td>50.7%</td>
</tr>
</tbody>
</table>

Table 1: Performance of LF Generation System

### 6 Recommendations

The sample data provided for the LFI task was used as the model for expected system behavior. A few recommendations that we believe would be improvements are below. Note that neither space nor time permitted an extensive consideration of other alternatives.

#### 6.1 Possessive Pronouns

Possessive pronouns are treated as mere adjectives in the sample data. A better way would be to handle these as any other possession indicator, and thus treat them as two-argument predicates.

Example:

*S: John drives his car.*

LF (Sample data):

\[
\text{John}:n_\text{(x1) drive}:v_\text{e1, x1, x2) his (x2) car}:n_\text{(x2)}
\]
6.2 Hybrid Verb Slot Representation

A verb's slots are supposed to signify their relation to the verb. The first slot is always reserved for the action/eventuality expressed by the verb. The second, then, is always for the subject of the verb. Now, the third slot should be reserved for the direct object of the verb (if any), and then, additional slots should be filled if and only if there are indirect objects associated with the verb.

We propose using either dummy or null arguments for certain slots. For instance, for a verb that has only indirect objects (apart from a subject), the representation for the verb can be, for instance,

\[
\text{run:}v_{(e1, x3, 0, x4, x6, ...)}, \text{or} \\
\text{run:}v_{(e1, x3, x9, x4, x6, ...)}, \text{where 'x9' is a dummy argument that is not referenced anywhere else in the Logic Form.}
\]

Moreover, the inclusion of prepositional attachments in the verb slots is a kind of redundancy that should be avoided. The following examples will make this proposal clear.

S: John plays at the park.
LF: John:n_{(x1)} play:v_{(e1, x1)} at (e1, x2) park:n_{(x2)}

S: John plays every day at the park.
LF: John:n_{(x1)} play:v_{(e1, x1, 0, x2)} every:a_{(x2)} day:n_{(x2)} at (e1, x3) park:n_{(x3)}

S: John plays tennis every day at the park.
LF: John:n_{(x1)} play:v_{(e1, x1, x2, x3)} tennis:n_{(x2)} every:a_{(x3)} day:n_{(x3)} at (e1, x4) park:n_{(x4)}

S: John gives Mary the book every evening in the library.
LF: John:n_{(x1)} give:v_{(e1, x1, x3, x2, x4)} Mary:n_{(x2)} book:n_{(x3)} every:a_{(x4)} evening:n_{(x4)} in (e1, x5) library:n_{(x5)}

Note that none of the examples include the arguments for ‘park’ or ‘library’ in the slots for verbs. Since these predicates are already connected to the verbs via prepositions, it is unnecessary to also include them in verb slots.

References


Altaf Mohammed. 2003. Logic Form Transformation of WordNet Glosses. Master’s dissertation, University of Texas at Dallas, Richardson, TX.


Logic Forms, particularly powerful logic representations presented in Moldovan and Rus (2001), are simple yet highly effective. In this paper, the structure of Logic Forms and their generation from input text are described. The results of an evaluation comparing the Logic Forms generated by hand with those generated automatically are also reported. SemEval (Semantic Evaluation) is an ongoing series of evaluations of computational semantic analysis systems; it evolved from the Senseval word sense evaluation series. The evaluations are intended to explore the nature of meaning in language. While meaning is intuitive to humans, transferring those intuitions to computational analysis has proved elusive. This series of evaluations is providing a mechanism to characterize in more precise terms exactly what is necessary to compute in meaning. As such...