Treebanks of Logical Forms: they are Useful Only if Consistent

Rodolfo Delmonte and Agata Rotondi

Department of Language Studies, Università Ca' Foscari - Ca' Bembo, 30123 – Venezia (It)
E-mail: delmont@unive.it Website: project.cgm.unive.it

Abstract

Logical Forms are an exceptionally important linguistic representation for highly demanding semantically related tasks like Question/ Answering and Text Understanding, but their automatic production at runtime is higly error-prone. The use of a tool like XWNet and other similar resources would be beneficial for all the NLP community, but not only. The problem is: Logical Forms are useful as long as they are consistent, otherwise they would be useless if not harmful. Like any other resource that aims at providing a meaning representation, LFs require a big effort in manual checking order to reduce the number of errors to the minimum acceptable – less than 1% - from any digital resource. As will be shown in detail in the paper, the available resources – XWNet, WN30-lfs, ILF - suffer from lack of a careful manual checking phase, and the number of errors is too high to make the resource usable as is. We classified mistakes by their syntactic or semantic type in order to facilitate a revision of the resource that we intend to do using regular expressions. We also commented extensively on semantic issues and on the best way to represent them in Logical Forms.

1. Introduction

In a number of recent papers, the need for a sizable (at least same size of WordNet) and publicly available corpus with Logical Form representation has increased: as a result more and more papers are concerned with the generation of a logical form or a semantic representation that is close to it. The fact is that there is already a number of such resources available, XWN (Moldovan and Rus, 2001), and ILF (Agerri and Peñas, 2010), hence (AP), both derived from WordNet glosses: so, why not using them. In fact in their paper, after reviewing previous work - including XWN and WN30-lfs (by Clark et al., 2008) generated by USC/ISI, California in 2006 - AP come to the conclusion that "... there is still some need for providing lexical and/or knowledge resources suitable for computational semantics tasks that required formalized knowledge." (ibid.29) The problem seems to be the presence of some obscurity in the way in which the glosses have been transformed - WN30-lfs is commented as containing "... free variables and/or predicates without any relation with any other predicates in the definition"(ibid.29) and the same problem is also present in XWN2 (ibid.,28). Here in addition, the output is cluttered with elements of the gloss which do not contribute to the definition strictly speaking, that is examples coming with the gloss. In fact also Clark et al. complain about the lack of consistency of XWN but no details are given.

Of course not all published comments on XWN speak negatively - without any detailed analysis, in fact - of XWN: on the contrary all published work by the authors of XWN speaks in favour of it. There are many papers published by the authors, V.Rus, D.Moldovan, S.Harabagiu et al., R.Mihalcea et al. – see the References -, who describe their work positively, if not highly positively, and comment on its usefulness for various semantically heavy tasks like Question Answering and RTE. In particular, Rus indicated an experiment with evaluation, where the accuracy for glosses conversion into Logical Forms is reported at 89.46%(Rus V., 2001), but on a selection of 1000 WN glosses only. The conclusion would be an error rate slightly over 10%, which is an important quantity of data but still perhaps bearable. In fact, we found over 30% error rate, and this is why – in our opinion - the XWN is badly flawed and cannot be used for the purpose it was made.

In the following sections we will go through the typical mistakes present in the corpus and comment on them. We don't want to imply that work carried out is useless, but since it can improved we intend to correct it in the future, and provinding classes of mistakes seems to best way to help doing that. A lot of difficult problems have been solved in XWN that deserve the resource to be saved and improved upon. Producing such a resource from scratch is outside the scope of current NLP technology, and this is attested by the various attempts at achieving such a goal (see also Ovchinnikova et al., 2011). However, there are also other attempts at producing Logical Forms directly from Penn Treebank style syntactic representations, like for instance, the LFToolkit by Nishit Rashod and Jerry Hobbs at their website, and the experiment reported by Alshawi et al. that we comment on here below.

In Alshawi et al. (2011) an experiment is reported to derive sentence-semantics pairs for training and testing from the Penn Treebank. In order to do that they program the Stanford treebank toolkit to produce what they call NLF expressions, that is Natural Logical Form, which are intentionally not intended as fully resolved logical forms. These are meant to be closer to natural logic than QLF Quasi Logical Forms, in order to use them to make some Natural Logic inference. And as the authors themselves comment, QLFs are being used widely to refer to any logic-like semantic representation without explicit quantifier scope, i.e. unscoped logical forms(ibid.17). In the same paper the authors specifically comment on the need to use an unknown/unspecified Null operator, %, for all those linguistic constructs which are beyond the

coverage of their semantic model. This applies to a great number of constructions that are present in the PTB and they give slightly different results in accuracy, both around 86%, however. Here again, we have to note that the usefulness of such logic-like representation is very low due to incompleteness of its results.

The Null operator is also present in PTB for all those linguistic constructions that have been regarded too difficult to take decisions upon by annotators and include all adjunct infinitivals and gerundives for a total amount of some 12,000 non coindexed null elements. This problem has also prevented other attempts at producing a semantically viable corpus of logical forms directly from a mapping of PTB, by a number of other researchers working in the LFG framework, (Guo et al.,2007) and in HPSG and CCG frameworks, but also Dependency Grammar as reported in (Nivre and Nilsson, 2005).

All these methods go beyond the encoding of surface context-free phrase structure trees, to incorporate non-local dependencies. This option requires recovering empty nodes and identifying their antecedents, be they traces or long distance dependencies. But since PTB annotators themselves intentionally refused to coindex all those cases that caused some difficulty in the decision process, all work carried out on this resource is flawed, semantically speaking, from the start. We must however, admit to the fact that WN glosses are much simpler sentences in comparison to PTB sentences, which even if taken with a word limit under 40 are still too complex and not comparable to definitions.

2. Common Mistakes and Their Classification

Logical Forms in XWN are graded in three quality levels: normal, silver and gold; the same applies to tagging and phrase structure constituency. "Normal" quality, which applies to the majority of the glosses, is used to indicate that there is no agreement between the two parsers that have been used to parse the input definition, and that there has been no manual checking of the output. "Gold" quality means manual checking has been performed, and "silver" quality indicates that there has been no manual checking but the two parsers agree in their representation. The importance given to the agreement between the two constituency parsers, is due to the fact that LFs are a mapping on syntactic constituency representation.

LF from glosses is represented in different manner according to lexical category, adjective, verb, noun and adverb: each one is associated to a predicate but with some differences. We list here below examples for each category:

A. Nouns.
For each synset the argument 'x1' is assigned to the first word that it represents. In the gloss, the 'x1' variable is referred to the same entity of the first word in the synset, as in:

plant:NN(x1) -> living:JJ(x1) organism:NN(x1) lack:VB(e1, x1, x2) power:NN(x2) of:IN(x2, x3) locomotion:NN(x3)

where *plant(x1)* and *living(x1) organism(x1)* refer to the same entity. One of the important efforts that characterizes positively XWN is the treatment of nominal compound, which has been done following Hobbs' suggestion in TACITUS to introduce the predicate NN in LF. Predicates may have a variable number of arguments and only the first is associated to the aggregation or compound of all the composing arguments, as in
jam_session:NN(x1) -> impromptu:JJ(x1) nn(x1, x2, x3) jazz:NN(x2) concert:NN(x3)

B. Verbs.
For each synset, the variable 'e1' is associated to the first term that represents it, to indicate the eventuality of the action/state/event of the verb meaning; the subject is associated invariably to 'x1' and the object to 'x2'. The second argument may be fictitious in case of intransitive verbs.

recognize:VB(e1, x1, x2) -> show:VB(e1, x1, x5) approval:NN(x3) or:CC(x5, x3, x4) appreciation:NN(x4) of:IN(x5, x2)

In this case all variables are bound to some argument position and are associated to some linguistic element. In the following example, an intransitive verb, we see on the contrary that there are two fictitious objects:

tremble:VB(e1, x1, x2) -> move:VB(e1, x1, x4) with:IN(e1, x3) tremor:NN(x3)

In the case of ditransitive verbs, the LF representation of the event is verb(e1,x1,x2,x3), as in, *professor gives students the grades: professor(x1) give(e1, x1, x2, x3) grade(x2) student (x3),* or in the definition of the verb GIVE:

give:VB(e1, x1, x2, x3) -> allow:VB(e1, x1, x3) to:IN(e1, e4) have:VB(e2, x3, x2) or:CC(e4, e2, e3) take:VB(e3, x3, x2)

C. Adjectives.
For each synset, argument 'x1' is associated to the first word that represents it, then in the second part of the gloss, argument 'x1' refers to the same entity described by the first word in the synset, as in,

ascetic:JJ(x1) -> practice:VB(e1, x1, x2) great:JJ(x2) self-denial:NN(x2)

D. Adverbs.

For each synset, argument 'e1' is assigned to the first term that represents it, then in the second part of the gloss, argument 'e1' refers to the same action, as in

grossly:RB(e1) -> in:IN(e1, x1) gross:JJ(x1) manner:NN(x1) largely:RB(e1) -> mainly:RB(e1) chiefly:RB(e1)

Other categories are treated as follows: prepositions are treated as predicates with two arguments, the first being the head noun that is modified by the prepositional phrase, and the second being the modified head noun; possessive pronouns introduce a relation between the governing head and the referent of the possessive pronoun, the predicate POS is used to represent this relation. What LFs do not contain are: verbal tense and mood (which can be regarded less relevant in definitions), negation, quantifiers (they are treated as adjectives or pronouns) and modal operators, comparative operators, plural, gender, illocutionary force and speech acts. Some of these semantic markers are only present, however, in few cases, as for instance in (A. Ramsay and D. Field, 2008).

We report here below common mistakes we found in the LF representation of XWN. This work has been carried out trying to group the most common mistakes into classes, be they related to tagging, to syntactic structure, to lexical types or semantic types. Of course for lack of space, we will not be able to discuss mor than one example per mistake. The first type of mistakes regards the disappearance of CONJUNCTIONS in coordinate structures and the consequent lack of binding of logical variables: here below we report the gloss focussing on the important portion of it and disregarding additional information.

Case 1: CONJUNCTIONS
Here, the missing conjunction is OR, and the unbound variable is "x5", also note that the coordinating conjunction AND is assigned variables which do not have any correspondence in the representation.

<gloss pos="NOUN" synsetID="07164600">
 <synonymSet>seedcake, seed_cake</synonymSet>
 <text> a sweet cake flavored with sesame or caraway seeds and lemon
 </text>
 <lft quality="NORMAL">
 seedcake:NN(x1) -> sweet:JJ(x1) cake:NN(x1) flavor:VB(e1, x7, x1) with:IN(e1, x6) sesame:NN(x2) caraway:JJ(x5) seed:NN(x3) and:CC(x30, x31, x32) lemon:NN(x4) </lft>
</gloss>

Case 2: PHRASAL VERBS AND PREPOSITIONAL VERBS
Most frequent prepositions appearing in the database are: on, in, to, by, for, with, at, of, from, as. Some of

them have an anomalous behaviour in the LF in case they appear at the end of the gloss: they are sometimes erased, and this does not depend on the quality of the LF because this happens in all three types, silver, gold or normal.

<gloss pos="VERB" synsetID="00042006">
<synonymSet>perfume, scent</synonymSet>
<text> apply perfume to; "She perfumes herself every day"
</text>
<lft quality="GOLD">
perfume:VB(e1, x1, x2) -> apply:VB(e1, x1, x3) perfume:NN(x3)
</lft> </gloss>

Here, on the contrary it is preserved,

<gloss pos="VERB" synsetID="00040699">
<synonymSet>powder</synonymSet>
<text> apply powder to; "She powdered her nose"; "The King wears a powdered wig"
</text>
<lft quality="GOLD">
powder:VB(e1, x1, x2) -> apply:VB(e1, x1, x3) powder:NN(x3) to:IN(e1, x2) </lft>
</gloss>

As to phrasal verbs the treatment is not homogeneous and sometimes the verb particle may simply be erased. It can appear attached to the verb as in (work_out:VB), or in a separate entry (set:VB(e1,x3,x1) up:IN(e1,x2)), or simply disappear,

<gloss pos="NOUN" synsetID="07918617">
 <synonymSet>secondary</synonymSet>
 <text>the defensive football players who line up behind the linemen
 </text>
 <lft quality="NORMAL">
 secondary:JJ(x4) -> defensive:JJ(x1) football:NN(x1) player:NN(x1) line:VB(e1, x1, x26) behind:IN(e1, x2) linemen:NN(x2) </lft>
</gloss>

Case 3: NOMINAL COMPOUNDS
As said above, nominal compounds are mapped into LF by means of the predicate *nn*. There is a great number of compunds which still have to be identified and mapped into LF, here however we refer to the case of a given compound which is identified but then it is mapped differently in different contexts. One such cases is constituted by "World War":

<lft quality="NORMAL">
bataan:NN(x1) -> peninsula:NN(x2) and:CC(x1, x2, x3) island:NN(x3) in:IN(x1, x4) philippines:NN(x4) japanese:JJ(x5)

force:NN(x5) besiege:VB(e1, x9, x5) american:NN(x6) force:NN(x7) in:IN(x6, x8) world_war_ii:NN(x8) </lft>

<lft quality="NORMAL">
wac:NN(x1) -> member:NN(x1) of:IN(x1, x2) women's:NN(x2) army:NN(x3) corp:NN(x4) be:VB(e1, x2, e2) organize:VB (e2, x9, x2) during:IN(e2, x5) world:NN(x5) war:NN(x6) ii:NN(x7) but:CC(e4, e0, e3) be:VB(e3, x1, x8) no:RB(e3) longer:RB(e3) separate:JJ(x8) branch:NN(x8) </lft>

<lft quality="SILVER">
battle_of_the_ardennes_bulge:NN(x1)   ->   battle:NN(x1) during:IN(x1, x2) world:NN(x2) war:JJ(x2) ii:NN(x3)
</lft>
<lft quality="NORMAL">
snafu:NN(x1) -> acronym:NN(x1) often:RB(e0) use:VB(e1, x2, x1) by:IN(e1, x2) soldier:NN(x2) in:IN(e1, x3) world:NN (x3) war:NN(x4) ii:JJ(x3) situation:NN(x5) normal:JJ(x6) all:JJ(x6) fucked:NN(x6) up:IN(e1, x6) </lft>

As can be noticed, the component words of "World War II" are analysed alternatively as separate Nouns, Nouns and Adjective, or as a single Noun.

Case 4: TAGGING ERRORS
The most frequent mistake in each of the four separate files is certainly the wrong POS assigned by the tagger. However in some cases the syntactic tree contains the right category while the LF has a corresponding wrong one.

<gloss pos="NOUN" synsetID="10317346">
<synonymSet>Hawking,   Stephen_Hawking, Stephen_William_Hawking</synonymSet>
<text> English theoretical physicist (born in 1942)
</text>
<lft quality="NORMAL">
hawk:VB(e1, x3) -> english:NN(x1) theoretical:JJ(x1) physicist:NN(x2)</lft>
</gloss>

<gloss   pos="ADV"   synsetID="00288722">
<synonymSet>clear, all_the_way</synonymSet>
<text> completely; "read ..."
</text>
<lft quality="GOLD">
clear:JJ(e1) -> completely:RB(e1)</lft>
</gloss>

Here the surname "Hawking" is turned into the verb "HAWK", then the ADVerbial "completely" is associated to an adjective JJ CLEAR. As will be commented below, there are many problems in the encoding of participles, as shown here again,

<gloss pos="NOUN" synsetID="00209984">
<synonymSet>chance-medley</synonymSet>
<text> unpremeditated killing of a human being in self defense
</text>

<lft quality="NORMAL">
chance-medley:NN(x1) -> unpremeditated:VB(e1, x5, x1) killing:NN(x1) of:IN(x1, x2) human:NN(x2) in:IN(x2, x3) self:NN(x3) defense:NN(x4) </lft>
</gloss>

and here,

<gloss   pos="NOUN"   synsetID="09420441">
<synonymSet>esthetician, aesthetician</synonymSet>
<text> a worker skilled in giving beauty treatments (manicures and facials etc.)
</text>
<lft quality="NORMAL">
esthetician:NN(x1) -> worker:NN(x1) skilled:VB(e1, x4, x1) in:IN(e1, e2) give:VB(e2, x1, x2) beauty:NN(x2) treatment:NN(x3) </lft>
</gloss>

Gerundives or present participles, when appearing at the beginning of a definition, are mapped onto the verb base form preceded by "act of", as in

advancing   toward   a   goal   ->   act:NN(x1)   of:IN(x1,   e1) advance:VB(e1, x2, x26) toward:IN(e1, e2) goal:NN(x2).

However, this should not happen when the –ing form is used as a nominalized verb as in

<gloss   pos="NOUN"   synsetID="05877558">
<synonymSet>notepaper</synonymSet>
<text> writing paper intended for writing short notes or letters
</text>
<lft quality="SILVER">
notepaper:NN(x1) -> act:NN(x1) of:IN(x1, e1) write:VB(e1, x2, x2) paper:NN(x2) intend:VB(e2, x6, x2) for:IN(e2, e3) write:VB(e3, x2, x5) short:JJ(x5) note:NN(x3) or:CC(x5, x3, x4) letters:NN(x4) </lft>
</gloss>

Genitive marking is interpreted in many different ways, as a Noun, Adjective or even Verb, in the Noun file, as shown here,

<gloss   pos="NOUN"   synsetID="00157666">
<synonymSet>capture</synonymSet>
<text>vthe removal of an opponent's piece from the chess board
</text>
<lft quality="NORMAL">
capture:NN(x1)   ->   removal:NN(x1)   of:IN(x1,   x2) opponent:NN(x2) 's:VB(e1, x2, x3) piece:NN(x3) from:IN(x3, x4) chess:NN(x4) board:NN(x5) </lft>
</gloss>

Case 5: FREE VARIABLES
Indexed variables are fundamental element of the LF and are used to indicate relations intervening between event and arguments or modifiers. In some cases, fictitious arguments can appear with free

variables at the event level, however when the argument is actually present - in particular, in intransitive or passivized structures -, it should be coindexed with the event. Very often this does not happen,

<lft quality="GOLD">
 hibernate:VB(e1, x1, x2) -> sleep:VB(e1, x1, x9) during:IN(e1, x3) winter:NN(x3) </lft>
</gloss>

<text> a man of such superior qualities that he seems like a deity to other people; "he was a god among men"
 </text>
<lft quality="NORMAL">
 god:NN(x1) -> man:NN(x1) of:IN(x1, x2) such:JJ(x2) superior:JJ(x2) quality:NN(x2) that:IN(e1, x5) seem:VB(e1, x2, x26) like:IN(e1, x3) deity:NN(x3) to:IN(x3, x4) other:JJ(x4) people:NN(x4) </lft>

<text> a commissioned officer in the United States Army or Air Force or Marines holding a rank above major and below colonel </text>
<lft quality="NORMAL">
lieutenant_colonel:NN(x1) -> commission:VB(e1, x11, x1) officer:NN(x2) in:IN(e1, x9) united_states_army:NN(x3) or:CC (x9, x3, x1, x4) air:NN(x1) force:NN(x4) or:CC(e3, e1) marine:NN(x5) hold:VB(e2, x5, x6) rank:NN(x6) above:IN(e2, x10) major:JJ(x8) below:IN(x20, x21) colonel:NN(x7) </lft>

Case 6. : NEGATION
There are lots of negations in WN glosses – 3107 cases of NOT overall - and as we will see, a number of them are wrongly scoped, some 20%. In particular, negation is distributed as follows in the four files: 2024 in Adjectives; 947 in Nouns; 79 in Adverbs; 57 in Verbs. If we add the other negation markers (NO, NONE, NOTHING, NEVER, NOR) adding up to 676 occurrences, we come up with some 3783 cases.
Negation can receive different scope according to its semantic role: it can negate the main verb or modifiers of the verb like adverbials – and in this case it will receive wide scope over the proposition, verb and arguments - or it can negate some specific argument or adjunct and in this case it will receive narrow scope. The majority of the cases of narrow scope negation is present in the Adjectival file: there are 901 cases of wide scope – that is the gloss is expressed by a full proposition with a verb and some argument; then there 1095 cases of narrow scope which is all correctly marked, as shown here below:

absolute:JJ(x1) -> not:RB(x1) limited:JJ(x1) by:IN(x1, x2) law:NN(x2)

Besides, consider the case of "alien" with the meaning of "foreign", where the negation has wide scope of the coordination of two verbs,

alien:JJ(x1) -> not:RB(e3) contain:VB(e1, x7, x1) in:IN(e1, x5) or:CC(e3, e1, e2) derive:VB(e2, x1) from:IN(e2, x2) essential:JJ(x2) nature:NN(x2) of:IN(x2, x3) something:NN(x3)

This is done extensively over all the dataset. Most errors derive from the wrong mapping of syntactic information in most of the case in which the negation is attached to an auxiliary verb, HAVE, BE, DO. In all these cases, the mapping wrongly produces two event variables, one for the auxiliary and another for the main verb, and the scope of negation is assigned narrow scope over the event variable of the auxiliary, as shown here below,

absentee_rate:NN(x1) -> percentage:NN(x1) of:IN(x1, x2) worker:NN(x2) do:VB(e1, x2, e2) not:RB(e1) report:VB(e2, x2, x26) to:IN(e2, e3) work:VB(e3, x2, x26)

However, in some cases the scope is marked correctly on the main verb as in,

lowbrow:JJ(x1) -> characteristic:JJ(x2) of:IN(x1, x2) person:NN(x2) be:VB(e1, x2) not:RB(x5) cultivated:JJ(x5) or:CC(e4, e1, e2) do:VB(e2, x2, e3) not:RB(e3) have:VB(e3, x2, x3) intellectual:JJ(x3) taste:NN(x3)

3   Some general considerations on XWN

Some general considerations over the whole dataset come from considering the amount of GOLD data with respect to NORMAL or SILVER, as shown in Table 1.

Types	Adverb.	Adjectiv.	Verbs	Nouns
Gold	3994	16059	14441	32844
Silver	0	4321	0	7228
Normal	0	0	0	54796
Total	3994	20380	14441	94868

Table 1.: Number of Gold/Silver/Normal LF entries in XWN

As can be easily gathered, the number of errors will vary substantially from one file to the other depending strictly on the number of GOLD LF entries, and will be proportional to the overall size of the file in terms of total number of entries. The file in which most errors are found is the one of NOUNS, which is not only the only file to contain Normal entries, but also in a quantity which is much higher than the GOLD ones, almost the double. Another important factor that may be considered as possible cause of errors in the NOUN file is the length of the gloss in number of words, which is more extended in syntactic terms than in the other files.
As a final remark, we extracted all the records containing just the LF from every single file, we then sorted them and checked for their consistency: this was

done in order to verify that no two Logical Form are identical to each other. Whenever this happens, the meaning associated to one synset would be interchangeable with the meaning associated to another synset, which is clearly a sign of inconsistency. We found the following situation,

- over 94868 entries for Nouns 43 are duplicate LFs
- over 20380 entries for Adjective, 47 are duplicate LFs
- over 3994 entries for Adverbs, 12 are duplicate LFs
- over 14441 entries for Verbs, 29 are duplicatre LFs

Here below we report some examples of duplicate, or sometimes triple LF representations taken from the Noun file:

alaska_peninsula:NN(x1) -> peninsula:NN(x1) in:IN(x1, x2) southwestern:JJ(x2) alaska:NN(x2)

alpaca:NN(x1) -> wool:NN(x1) of:IN(x1, x2) alpaca:NN(x2)

anagoge:NN(x1) -> mystical:JJ(x1) allegorical:JJ(x1) interpretation:NN(x1)

approbation:NN(x1) -> official:JJ(x1) approval:NN(x1)

bailey:NN(x1) -> outer:JJ(x1) courtyard:NN(x1) of:IN(x1, x2) castle:NN(x2)

Bernoulli:NN(x1) -> swiss:JJ(x1) mathematician:NN(x1)

blood_count:NN(x1) -> number:NN(x1) of:IN(x1, x2) red:JJ(x2) white:JJ(x2) corpuscle:NN(x2) in:IN(x2, x3) blood:NN(x3) sample:NN(x4)

card_catalog:NN(x1) -> enumeration:NN(x1) of:IN(x1, x2) resource:NN(x2) of:IN(x2, x3) library:NN(x3)

cassava:NN(x1) -> source:NN(x1) of:IN(x1, x2) tapioca:NN(x2)

catapult:NN(x1) -> use:VB(e1, x2, x1) to:IN(e1, e2) propel:VB(e2, x1, x1) small:JJ(x1) stone:NN(x1)

clash:NN(x1) -> state:NN(x1) of:IN(x1, x2) conflict:NN(x2) between:IN(x2, x3) person:NN(x3)

4. Intermediate Logical Forms

In their paper (Agirre & Peñas, 2010) the authors present an automatic system that produces LFs from WordNet glosses using Stanford Parser and then mapping the output with typed dependencies into what they call ILFs. The important contribution of these two authors is the preprocessing phase of the glosses in order to make them concise and homogeneous as much as possible. To this aim, the authors eliminate all content between parenthesis; they also eliminate all that comes after a semicolon. Then they treat the three main categories as

follows: they add a period at the end of the gloss for all categories; nouns and adverbs have the first word capitalized; in the case of adjectives, they add the word "Something" at the beginning of the gloss; and with verbs, they add the particle "To" at the beginning.

If we compare the result obtained in ILF with the LFs of XWN we notice that the mistakes that we found and commented above still occur but with a much lower frequency. The most important mistake we noticed in XWN here it is totally absent: there are no unbound variables in LF, all the variables are bound regularly. In addition to XWN ILF contains all article, conjunctions and prepositions.

Overall, we have noticed a remarkable improvement in the LF representation but as the authors themselves comment, the resource needs improvement. In particular there is no word sense assigned uniquely to each gloss as happened in XWN. This could be easily amended given the availability of a newly released version of the glosses with sense disambiguation at WordNet website. The resource still needs some comprehensive evaluation and, as the authors themselves indicate, this will be done when version 1.0 will be available.

The first observation to be made is that the reduction and also the fact that the authors managed to focus on the definition and eliminated most if not all of the remaining additional unessential parts, is certainly to be judged positively. However, as we show below, the resulting Logical Form has on the contrary become less readable if not unreadable and difficult to use, in one word it has lost perspicuity. Consider one example:

<sense offset="301890382" pos="s" synset_name="bigheaded.s.01"> <gloss> <text>Something overly conceited or arrogant.</text> <parse parser="Stanford parser 1.6.1">
....
<ilf version="0.2">[rel(1,3,2,'advmod',G1_3,G1_2), rel(1,1,3,'amod',G1_1,G1_3), rel(1,1,5,'amod',G1_1,G1_5), rel(1,3,5,'conj_or',G1_3,G1_5), e(1,2,G1_2), w(1,2,'overly','r','rb'), e(1,3,G1_3), w(1,3,'conceited','a','jj'), syn(1,3,301891773), e(1,1,G1_1), w(1,1,'something','n','nn'), e(1,5,G1_5), w(1,5,'arrogant','a','jj'), syn(1,5,301889819)]</ilf>
<pretty-ilf>something(x1) amod(x1,x3) amod(x1,x5) overly(x2) conceited(x3) advmod(x3,x2) conj_or(x3,x5) arrogant(x5) </pretty-ilf>

The authors have cluttered the LF with all details derived from the dependency graph produced by Stanford's parser, including tags associated to words, dependency types, which are rendered as rel(ations) on the arc linking two words. Then words are eventually associated to their dependency indices but in addition, they have double tags, the ones produced by their tagger and the ones coming from Stanford's parser. Eventually they provide a "pretty print" version of ILF – for "easier readability" as they say (ibid.,33) – where a straightforward version appears with just words and variables. Here they use dependency types

as prefixes: the net result is that the LF is populated by eight expressions just like what the original complete gloss would require. Here below we report the complete version of the LF produced in XWN, where the adjectives are simply treated as modifiers of the same head and the conjunction is erased,

<entry word="bigheaded#a#1" status="partial">
<gloss>used colloquially of one who is overly conceited or arrogant</gloss>
bigheaded:JJ(x1) -> use:VB(e1, x6, x1) colloquially:RB(e2)
of:IN(e1, e2) one:JJ(x3) be:VB(e2, x1) overly:RB(x4)
conceited:JJ(x4) arrogant:JJ(x4)

As can be noticed, ILF has reduced the linguistic content of the gloss but in so doing it has deleted important information regarding the register of usage of the main entry word marked as "colloquial". Also, the introduction of dependency types has made the overall LF representation less perspicuous and certainly difficult to use in practical applications.
Mistakes we found are as follows:
- not all contents within parenthesis have been eliminated:

<text>Pure ethyl alcohol (containing no more than 1% water).</text>
    <pretty-ilf>pure(x1) ethyl(x2) alcohol(x3) amod(x3,x1)
nn(x3,x2) ((x4) nsubj(x4,x3) dep(x4,x5) contain(x5)
advmod(x5,x6) dobj(x5,x12) no(x6) dep(x6,x10) more(x7)
than(x8) advmod(x8,x7) 1(x9) quantmod(x9,x8) %(x10)
num(x10,x9) water(x11))(x12) nn(x12,x11)</pretty-ilf>

The use of the predicate "nn" for compound nouns has been improved and we checked that also for the compound commented above, "World War II", which is mapped correctly; this notwithstanding there are many unneeded uses of the predicate "nn" as for instance in "coarse tobacco",

    <text>A strong coarse tobacco that has been shredded.</text>
<word ind="1" pos="DT">a</word>
    <word ind="2" pos="JJ">strong</word>
    <word ind="3" pos="NN">coarse</word>
    <word ind="4" pos="NN">tobacco</word>
    <pretty-ilf>a(x1) strong(x2) coarse(x3) tobacco(x4)
det(x4,x1) amod(x4,x2) nn(x4,x3) rcmod(x4,x8) that(x5)
have(x6) be(x7) shred(x8) nsubjpass(x8,x4) rel(x8,x5)
aux(x8,x6) auxpass(x8,x7)</pretty-ilf>

where we assume that there might have been a tagging error. More tagging errors occur with colour nouns and past participles. Other mistakes come from wrong cases of pp_attachment as for instance in the following entry,

<text>The nonrandom movement of an atom or radical from one place to another within a molecule.</text>
    <pretty-ilf>the(x1) nonrandom(x2) movement(x3)
det(x3,x1) amod(x3,x2) prep_of(x3,x6) prep_of(x3,x11)

prep_to(x3,x13) a(x5) atom(x6) det(x6,x5) conj_or(x6,x11)
radical(x8) prep_from(x8,x10) one(x10) place(x11)
amod(x11,x8) another(x13) prep_within(x13,x16) a(x15)
molecule(x16) det(x16,x15)</pretty-ilf>

Maybe the mistake here is caused by the wrong tag associated to RADICAL which is treated as JJ rather than as NN. It is obvious that by using Stanford parser a certain level of error rate is expected: it would have been interesting to know what additional error rate is introduced by the conversion algorithm, but the evaluation is missing yet. It is also important to remember that Stanford parser only produces a surface level representation with some additional predicate argument completion for passive structures and some control infinitivals. So it is impossible to judge whether the reduction process – also in light of the example discussed above – has positively contributed to the final representation or not. Certainly the most important contribution, the elimination of free variables and the control exerted on the predicates arity, constitute by themselves already an important goal achieved. Of no real consequences is on the contrary the added feature regarding the insertion of the sense synset index directly in the overall logical form representation, the one delimited by ILF: this fact is disputable simply by iteself seen that there has been no word sense disambiguation of the gloss as a whole, something commented upon also by the authors in their conclusions (ibid.,35).

5. Conclusions

Eventually we may comment that there are a number of resources available with Logical Forms representations of WordNet glosses, and a number of algorithms which can be used off-the-shelf to produce Logical Forms from PTB constituency based phrase structure representations: none of these resources is however usable as is, do to error rates which average 30%. Improvements can be achieved by manual correction of all the LFs contained in these resources. This is an option that we intend to carry out in a local project that will be the followup of a MA degree thesis that started this research. The research has focussed on the typing of the mistakes present in the resource itself: this has been made easier by the fact that in both resources analysed, the conversion into LFs has started from the output of a syntactic parser – in the case of XWN, two constituency parsers, while in ILF, one dependency parser. The result of the manual corrections will be made available online to be accessed freely by anyone interested in using them.

References

Agerri, R. and Anselmo Peñas (2010). On the Automatic Generation of Intermediate Logic Form for WordNet glosses, In *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (Cicling-2010),* LNCS Vol.

6008, Springer.

Alshawi, H., Pi-Chuan Chang, M. Ringgaard. (2011). Deterministic Statistical Mapping of Sentences to Underspecified Semantics, in Johan Bos and Stephen Pulman (editors), Proceedings of the 9th International Conference on Computational Semantics, IWCS,15-24.

Clark, P., Fellbaum, C., Hobbs, J.R., Harrison, P., Murray, W.R., Thompson, J.: Augmenting WordNet for Deep Understanding of Text. In: Bos, J., Delmonte, R. (eds.) *Proceedings of STEP 2008, Conference on Semantics in Text Processing. Research in Computational Semantics,* vol. 1, pp. 45–57. College Publications (2008)

Delmonte R., (2007). Computational Linguistic Text Processing – Logical Form, Semantic Interpretation, Discourse Relations and Question Answering, Nova Science Publishers, New York.

Yuqing Guo, HaifengWang, Josef van Genabith, Recovering Non-Local Dependencies for Chinese, 2007. In *Proceedings of the Joint Conference on EMNLP and Computational Natural Language Learning,* pp. 257–266.

Harabagiu, S.M., Miller, G.A., Moldovan, D.I.: *eXtended WordNet - A Morphologically and Semantically Enhanced Resourc*e (2003).
    Available at http://xwn.hlt.utdallas.edu.

Information Science Institute, University of Southern California: Logical Forms for WordNet 3.0 glosses (2007), http://wordnetcode.princeton.edu/standoff-files/wn30-lfs.zip

Mihalcea, R., and Dan I. Moldovan, (2001). *eXtended WordNet: progress report,* In: Proceedings of NAACL Workshop on WordNet and Other Lexical Resources, Pittsburgh, 95-100.

Moldovan, D., Rus, V.: Explaining Answers with Extended WordNet. In: Proceedings of the Association for Computational Linguistics, ACL 2001 (2001)

Moldovan, D., S. Harabagiu, R. Girju, P. Morarescu, F. Lacatusu, A. Novischi, A. Badulescu, O. Bolohan, *LCC Tools for Question Answering,* In: Proceedings of TREC, 2002.

Moldovan, D., Vasile Rus: Transformation of WordNet Glosses into Logic Forms. FLAIRS Conference 2001: 459-463.

Moldovan, D., Vasile Rus: Logic Form Transformation of WordNet and its Applicability to Question Answering. ACL 2001: 394-401.

Ovchinnikova, E., N. Montazeri, T. Alexandrov, J. R. Hobbs, M. C. McCord, R. Mulkar-Mehta, (2011). Abductive Reasoning with a Large Knowledge Base for Discourse Processing, in J. Bos and S. Pulman (editors), In Proceedings of the 9th International Conference on Computational Semantics, IWCS, 2Oxford, pp. 25-234.

Ramsay, A.M. and D.G.Field (2008). Speech acts, epistemic planning and Grice's maxim. Journal of Logic and Computation 18(3), 431-457.

Rus, V. and Dan I. Moldovan: High Performance Logic Form Transformation. International Journal on Artificial Intelligence Tools 11(3): 437-454 (2002)

Rus, V., Dan I. Moldovan, Orest Bolohan, (2002). Bracketing Compound Nouns for Logic Form Derivation. FLAIRS Conference 2002: 198-202.

Rus, V., Alex Fit-Florea, (2004). A Model for Identifying the Underlying Logical Structure of Natural Language. PRICAI 2004: 977-978.

Rus, V., (2004). Experiments with Machine Learning for Logic Arguments Identification. MAICS 2004: 40-47.

Rus, Vasile, (2004). *A First Evaluation of Logic Form Identification Systems,* SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text.

Vasile Rus, *High Precision Logic Form Transformation,* 2001. In Proceedings of ICTAI'2001. pp.288-288.

WordNet Gloss Disambiguation Project, Princeton University: Semantically annotated gloss corpus (2008), http://wordnet.princeton. edu/ glosstag.shtml