Evaluation of the XIP German Grammar

(september 2006, S. Maurel)

1. Introduction

The purpose of this document is to comment the results of the evaluation of the XIP German grammar (by september 2006).

The whole evaluation is in the file **results_eval.xls** in the repertory EVAL. The following tables (the values are rounded to two decimal places) resume the result. In fact, there are two evaluations, the first one is based on the disambiguation with the HMM-tagger (sheet 1), and the second one with a disambiguation without HMM (sheet 2).

Seven texts (containing 433 sentences, 6067 words) were analyzed with the general German grammar. The utilized disambiguation, chunking and dependencies files were copied and renamed in the repertory EVAL with the suffix **_eval1.xip** for the first evaluation, and **_eval2.xip** for the second evaluation (the chunking and dependencies files of the second evaluation are only adapted as far as necessary on the new disambiguation file).

The texts are newspaper articles from different newspapers and coming from different resorts. They were taken from the internet around mid-august. The last text is a sample of 100 sentences of the Pocket Engineer Corpus which contains technical instructions about printers. (This text didn't change as much the results as we expected, the sample of 100 sentences seems to contain mainly well-formed sentences.)

The *Precision*, *Recall* and *F-Score* were calculated for the following dependencies:

- **THEMA** (the "subject" of a verbless sentence, an unary relation),
- **SUBJ** (subject of the verb, binary relation),
- **OBJA / OBJ_ACC** (accusative (direct) object, binary relation),
- **OBJD / OBJ_DAT** (dative (indirect) object, binary relation),
- **OBJG / OBJ_GEN** (genitive object, binary relation)
- and **OBJS / OBJ_SENT** (sub clause object, binary relation between the verb and the conjunction that introduces the subordinate clause).

The *Precision* corresponds to the quotient of the sum of the true positive results and the sum of the true positives plus the false positives. The *Recall* corresponds to the quotient of the sum of the true positive results and the sum of the true positives plus the false negatives. And the *F-Score* is a combination of the two previous values; it is the quotient of the product of 2 times the precision and the recall, and the sum of precision and recall (more details in results_eval.xls).

True positive are the analyses which extract correctly the dependency relations, false positive are the analyses which extract a dependency relation where this relation is not present in the text and false negative are the analyses which didn't extract the existing dependency.
Right of each dependency (in results_eval.xls) is a colon which lists the possible reasons for the erroneous analyzes.

2. First evaluation (with HMM)

	Subject	Objects	Thema	All dependencies
Precision	0.86	0.5	0.86	**0.74**
Recall	0.71	0.56	0.99	**0.75**
F-Score	0.78	0.51	0.91	**0.73**

The most problems are due to not sufficiently disambiguated nouns which have features of several cases (nominative, genitive, dative and accusative), and ambiguities between verbs and adjectives, finite and infinite verbs, and conjunctions and prepositions.

The genitive object didn't appear at all in the text sample, but was extracted many times. It seems that this object dependency is quite rare and should be restricted to some special verbs only, e.g. with a special feature.

Often the sentences were to long and to complex and didn't fit the context that defines the rule, and so the dependencies weren't extracted or only partially.

There should be special rules for named entities like persons, dates, locations, etc. The quotes around names of newspapers presented a problem quite often, also if there was direct speech in the sentence.

Very problematic was the coordination of more than two elements. Or when the second element of a PP-coordination wasn't in that PP, it could provoke erroneous object dependencies.

Numbers should be chunked with the following word, if it is "Jahr", a currency or "Prozent", etc. and some multiwords should be added in the lexicon or the local grammar, like "Studium Generale".

The new orthography allows the adverb "heute abend" to be written as "heute Abend", this should also be coded as a multiword, otherwise the word "Abend" can become subject when it is chunked as a normal single noun.

The word order in the sentence follows not always the SVO (subject, verb, object) schema, so the rule that deletes the subject following the verb if there is another potential subject before the verb (the object in some cases) should be reviewed.

3. Second evaluation (without HMM)

	Subject	Objects	Thema	All dependencies
Precision	0.82	0.43	0.78	**0.68**
Recall	0.74	0.57	0.93	**0.75**
F-Score	0.78	0.47	0.83	**0.69**

Unfortunately, the most problems are still due to not enough disambiguated nouns and verbs. Especially verbs were not correctly disambiguated between finite verbs, infinitives and adjectives. Therefore the corresponding subjects and objects haven't been extracted correctly. In general there were to much finite verbs (that are past participles, infinitives and adjectives in reality) which were falsely combined with the nouns to subject or object dependencies.

The Pocket Engineer text contains a lot of sentences which start with a nominalized verb which is tagged in this grammar as a finite verb. So the THEMA dependency cannot be established and often a wrong SUBJ relation is extracted.

There were still some genitive objects extracted, but not so much as in the first version of the grammar thanks to the better case disambiguation of the nouns. Some dative objects were extracted where there should be only an accusative object.

Like in the previous version, names of newspapers and direct speech that appear in quotes are problematic for the system, also product names which are made of more than one word.

For the rest, it was mainly the same problems as for the first version ("heute Abend", "Studium Generale", the word order that differs from SVO, etc.).