

# The performance of a grammar checker with deviant language input

Janne Bondi Johannessen, Kristin Hagen and Pia Lane  
The Text Laboratory, University of Oslo  
Pb 1102 Blindern  
0317 Oslo, Norway  
{j.b.johannessen, kristin.hagen, p.m.j.lane}@ilf.uio.no

## Abstract

This paper reports on an evaluation performed on the Grammar Checker for Norwegian (NGC), developed at The Text Laboratory, University of Oslo.<sup>1</sup> The ability of the NGC to find errors made by different “non-standard” linguistic groups is analysed and compared to its performance when tested on texts written by “standard” users. Then possible ways of adapting the NGC for use on deviant language input are discussed.

## 1. Introduction

This paper reports on the results of an evaluation we have performed on the Grammar Checker for Norwegian (NGC), developed at The Text Laboratory, University of Oslo. The NGC is now part of Microsoft Word in the Office XP package released in 2001.<sup>2</sup> The goal of the NGC was decided partly by that of the Swedish Grammar Checker (SGC, Arppe 2000 and Birn 2000), designed to detect what were assumed to be the errors of “standard” users, and partly by a wish to include more linguistically advanced features. The kind of grammatical mistakes made by linguistically “non-standard”<sup>3</sup> groups was not taken into account, and this kind of tool obviously would be beneficial to these groups.

Having provided an overview of the main method behind the NGC, we will give a general overview of the kinds of errors that the NGC is designed to detect. Then we will show how it performs on various deviant language input

(essays written by Slav and Chinese students, and Norwegian deaf children).

## 2. The features of the NGC

### 2.1 The method behind the NGC

The NGC was developed using Constraint Grammar (Karlsson et al. 1995). Like the SGC the NGC has three main parts in addition to an initial tokenizer (spell checking is performed at a previous stage):

- A morphological analyser (NOBTWOL), which provides each word form with all of its lexically possible readings (grammatical tags).
- A morphological CG disambiguator, which eliminates incorrect tags according to the grammatical context (Karlsson et. al 1995, Hagen, Johannessen and Nøklestad 2000a and 2000b).
- An error detector that identifies different kinds of grammatical errors.

There is an interesting problem regarding the construction of a grammar checker. On the one hand it is necessary to have as much grammatical information as possible about the particular text that is going to be checked. On the other hand, it is very difficult to perform any such grammatical analysis, since grammatical features (“errors”) essential for the analysis might be missing. We tried to solve the problem by relaxing many of the requirements of the disambiguating tagger described above, since it was originally developed for grammatically correct texts. An example of this is the original CG rule assigning a determiner reading to a word that is next to a noun and agrees with it in number and gender:

---

<sup>1</sup> <http://www.hf.uio.no/tekstlab/>

<sup>2</sup> The NGC was developed for the Finnish company Lingsoft <http://www.lingsoft.fi/>.

<sup>3</sup> Non-native speakers, deaf people, aphasics and dyslexics.

- (01) (@w =! (det neut)  
 (0 DEF-DET)  
 (\*1 DEF-SG-NEUT-NOUN \*L)  
 (NOT LR0 NOT-ADJ-NOUN \*L)  
 (NOT \*L NOT-ADV-ADJ))

The rule (one of approximately 2000 rules) says that if a word is definite and has neuter determiner as one of its readings, but there is a neuter definite singular noun to its right, with nothing but adverbs and adjectives in between, then the determiner reading is correct. This rule ensures that the first word in the sentence below is correctly tagged as a determiner and not e.g. a pronoun:

- (02)  
*Det eplet likte han godt*  
 the.DEF.NEUTER.SG apple.DEF.NEUTER.SG liked he well  
 'That apple, he liked well.'

The tagger can then safely assume that whatever does not agree with the noun to its right is not part of the same noun phrase, and therefore is a pronoun. However, a *grammar checker* can never assume that anything is correct, and cannot rely on the agreement features of the determiner and the noun. Instead, it ought to be able to detect any missing agreement and point out the error. So the new relaxed tagger leaves more ambiguity. Instead, very specific error rules are introduced in the NGC. Rule (03) below (one of 700 error rules) detects gender disagreement between a determiner and the following noun (04).

- (03) (@w =s! (@ERR)  
 (0 DET-DEF-NEUT)  
 (NOT -1 DITRANS)  
 (1C NOUN-SG-DEF)  
 (NOT 1 NEUT)  
 (1 MASC))

- (04) \*Jenta så det bilen  
 The.girl saw the.DEF.NEUT.SG car.DEF.MASC.SG  
 'The girl saw that car.'

This method is reminiscent of that suggested by Schneider and McCoy (1998) for their ICICLE system designed to help second-language learners of English. However, since theirs is a grammar based on context-free rules, it is more

difficult to implement; in order for a parse to be successful, all phrases have to be well-formed, which means that the grammar must include rules for ungrammatical structures. CG has an advantage; it does not have to build a full phrase structure, thus partial parses are fine, and local errors are easily detected.

## 2.2 Error types

The NGC detects the following main error types:

- Noun phrase internal agreement:
  - Definiteness *et huset --> et hus*  
*a the.house --> ... house*
  - Gender agreement *en hus --> et hus*  
*a.MASC. house.NEUT --> a. NEUT ...*
  - Number agreement *et epler --> et eple*  
*an apples --> ... apple*
- Subject complement agreement  
*Bilen er rød --> Bilen er rød*  
*The car.MASC is red.NEUT.--> ... red. MASC*
- Negative polarity items  
*Jeg kjøper noen bok --> Jeg kjøper ei bok*  
*I buy any book --> ... a book*
- og/å errors (conjunction/ inf. marker)  
*De gikk å sang --> De gikk og sang*  
*They walked to sang --> ... and sang*
- Too many or no finite verb(s) in a sentence  
*I Norge er var det slik --> I Norge er det slik.*  
*In Norway is was it such --> ... is it such*
- Word order errors  
*Nå gutten kom --> Nå kom gutten.*  
*Now the.boy came --> Now came ...*

## 2.3 Previous evaluation of the NGC

Our guide line, given to us by Lingsoft, for the acceptable number of "false alarms" was 30% (70% of all alarms had to report true errors), and it performs well within that limit, with a precision of 75% (Hagen, Johannessen and Lane 2001), compared with 70% for the SGC (Birn 2000). The recall rate for the NGC has not been calculated.

The figures above were calculated on the basis of texts written by advanced language users - mostly Norwegian and Swedish journalists, with few errors in each text. Most of the errors were not due to lack of knowledge of Norwegian grammar, but rather to modern word processing: too quick use of functions like cut and paste, insert etc. For example, two finite

modal verbs next to each other would not be uncommon. However, one would assume that less linguistically advanced users might benefit more from this kind of tool. In the next sections we shall evaluate the NGC on texts produced by various non-standard language users.

### 3. New evaluation

#### 3.1 Errors and non-standard language users

We have so far tested four groups of foreign students and one group of Norwegian deaf pupils, and are in the process of testing aphasics and dyslexics. We have divided the errors into five groups:

**i) Idiomatic errors:** This covers language use not strictly speaking ungrammatical, just «foreign», **ii) Lexical errors:** Wrong word, lack of subcategorised word, or a word too many, **iii) Syntactic errors:** Wrong word order, lack of word (that's not subcategorised by a particular word), negative polarity errors, wrong choice of pronoun/anaphor, **iv) Morphological errors:** Morphological features, NP agreement (number, definiteness, gender), predicative agreement, tense of verbs, **v) Pragmatic errors:** Errors that involve sentence-external rules: Definiteness of NPs (due to known or new information), verb tense that ought to follow from the context.

More specifically, we have tested the NGC on essays written by Norwegian deaf pupils (11-15 years old) and four groups of foreign university students in Norway (Slav and Chinese students on Level II (Intermediate) and Level III (Advanced)). We have included papers written by a control group of Norwegian pupils, as the student essays were hand written and the initial precision of the NGC was calculated on word-processed texts. We will also test the NGC on essays written by dyslexic and aphasic adults.

#### 3.2 Test results

There is not enough space to give the individual test results here. Let us instead illustrate with one group, the Chinese intermediate students. There were 15 essays of an average of 300 words, altogether 4500 words, the same amount

as for the other test groups. The vast majority of the detected errors are morphological ones, see table (05):

(05) Errors detected by the NGC for Chinese Level II stud.

Error types	Number
Syntactic	4
Morphological	28

(06) *Example of syntactic error:*

Fordi jeg kan ikke uttrykke meg  
because I can not express myself  
⇒ Fordi jeg ikke kan uttrykke meg

(07) *Example of morphological error:*

Taiwan er et lite øy  
Taiwan is a (neut) small (neut)  
island (masc)  
⇒ Taiwan er en liten øy

However, in order to evaluate the NGC properly with respect to the Chinese students, we have to look at all errors made.

(08) Errors by Chinese Lev. II stud. not found by the NGC :

Error types	Number
Syntactic	68
Morphological	45
Lexical	70
Pragmatic	13
Idiomatic	32
<b>Sum:</b>	<b>228</b>

In addition to the 32 errors detected by the NGC, the Chinese Level II students made 228 errors that were not detected by the NGC, i.e. only 12% were found. But notice that nearly half the errors (115) are lexical, idiomatic and pragmatic ones – error types that have not even been attempted to be detected by the NGC.

(09) *Example of idiomatic error:*

Nå er jeg i Norge som alle er dyre  
now am I in Norway which all are expensive (pl)  
⇒ Nå er jeg i Norge hvor alt er dyrt

(10) *Example of lexical error:*

Jeg var veldig redd av blod  
I was very afraid of blood  
⇒ Jeg var veldig redd for blod

(11) *Example of pragmatic error:*

Det er en vane du må etablere når du var barn  
It's a habit you must establish when you were child  
⇒ Det er en vane du må etablere når du er barn

Of the morphological mistakes made by the Chinese Level II students, the NGC detected 28 out of 73, a recall of 38% - considerably higher than the results for all categories taken together. It can also be improved by adding more morphological rules.

This is similar to the error pattern of all the other non-standard language groups we have studied so far (Chinese Level III students, two levels of Slav students and deaf Norwegian pupils). The NGC finds 10% of the total number of errors in the essays written by Slav students. For the deaf students, the NGC findings rise slightly, to 14%. A reason for the higher percentage could be that the deaf pupils make many morphological mistakes, a feature the NGC is designed to detect. For example, these pupils typically use non-finite verb forms and wrong gender for nouns.

Like the Chinese students, both the Slavs and the deaf pupils have a very high percentage of «non-grammatical» errors, i.e., lexical, idiomatic and pragmatic. The non-grammatical errors of the Slav students amount to 60% of all errors, while the number for the deaf pupils is 52%.

However, there are also big differences between the groups, see table (12) below. For example, the foreign language students have fewer idiomatic and pragmatic errors than the deaf pupils (20% of all errors versus 31%). This aspect is even more striking when we look at the pragmatic errors only. The Slav students have only 4% pragmatic errors (of all errors). The Chinese students have a higher number; 9%. The deaf students, however, have 22% pragmatic errors.

(12) Errors in % of all errors

Error types	Chinese	Slav	Deaf
Syntactic	23	17	15
Morphological	24	23	37
Lexical	31	41	17
Pragmatic	9	4	22
Idiomatic	12	15	9

The deaf students especially make two kinds of pragmatic errors: wrong choice of definiteness on the basis of given/new

information, and wrong use of tense (typically a change of tense when none is called for). Related to this is the morphological kind of error mentioned above: lack of finiteness on verbs. These numbers, though interesting, are hardly surprising; to some extent they reflect the linguistic background of these language users. The Norwegian Sign Language and Chinese have no morphological verb marking or noun marking, while Slavic languages have a complex system of verb inflection.

The results for the Norwegian control group are predictable. They make no non-grammatical mistakes, few grammatical mistakes<sup>4</sup>, and frequently split compounds incorrectly. 16% of their errors were found by the NGC – slightly higher than the number for the other test groups, but much lower than the equivalent number of the SGC which was reported to be 35% (Birn 2000) in Swedish newspaper texts. Obviously, the reason for the lower number is that the essays by the Norwegian pupils are originally written by hand, and thus lack easily detectable cut-and-paste and our word-processing errors. Our ongoing research will show us the results for the other "non-standard" language groups.

The NGC gives surprisingly few «false alarms» (the precision is 95%, as opposed to 75% for the newspaper texts) in the texts by non-standard language groups, due to the fact that their language is very simple, suiting the shallow analysis performed by the NGC. The precision for the Norwegian control group is also high: 87%.

#### 4 Can the NGC be improved?

With a larger-scale error analysis of authentic texts from the non-standard groups a lot of new knowledge could be found, which would make a good basis for improving the NGC. More specifically, since morphological and syntactic features are governed by sentence-internal rules, a rule-based grammar checker like the NGC

<sup>4</sup> Apart from *og/å* errors (conjunction and inf.marker – notoriously difficult because the pronunciation is the same)

should be able to account for violations of such features.

In fact, we have done some testing for this purpose, and developed rules for 14 new morphological and syntactic error types<sup>5</sup>. One new feature is the detection of compound words erroneously written as two separate words (*kreft studier* 'cancer studies' instead of *kreftstudier*). The results are promising: The overall recall for morphological errors for our Chinese students now increased from 38% to 53%.

Detection of non-grammatical errors is more difficult to improve. Of course, with an advanced lexicon containing e.g. detailed subcategorisation information, we might be able to find some lexical errors, such as certain prepositions after certain verbs. On the other hand, a lexicon instead of context, since something which looks like a subcategorised category may in fact be something else. Idiomatic expressions are more of a problem, but given a large error corpus, it would be possible to extract the most common error types and look for these later. Pragmatic errors are notoriously difficult. But even here, it would be possible to use some system to alarm the user whenever, for example, a singular count noun occurs without a determiner or two verbs have different tense in the same sentence.

### Conclusion

The Norwegian Grammar Checker was made for native Norwegian speakers. We have tested the NGC on texts written by "non-standard" language users (Slav and Chinese students of Norwegian, and Norwegian native deaf pupils). It turned out that a high percentage, as much as 85-90%, of their mistakes remained undetected. However, the picture is not quite as bleak as it might seem, since the majority of the errors, 50-60%, were "non-grammatical", i.e. lexical, idiomatic and pragmatic, errors which the NGC was not designed to detect.

Does this mean that a grammar checker is of no help for these groups? We do not think so; 40-50% of the errors were grammatical, i.e. subject

to sentence-internal linguistic rules. Since the NGC is rule-based, it should be possible to account for many of these kinds of errors. Indeed, preliminary tests indicate that this is the case.

### Acknowledgements

Our thanks go to Wenche Vagle (the KAL project) for providing Norwegian essays, The Department of Norwegian for Foreign Students at the UiO, for texts from foreign students, Elisabeth Svinndal and Skådalen skole for texts from deaf pupils, and to Pål Kristian Eriksen for various practical and linguistic help.

### References

- Arppe, A. 2000. Developing a grammar checker for Swedish. In Nordgård, T. (ed.) *Nodalida '99 Proceedings from the 12<sup>th</sup> Nordiske datalingvistikkdager*, Department of Linguistics, University of Trondheim, p. 13-27.
- Birn, J. 2000. Detecting grammar errors with Lingsoft's Swedish grammar checker. In Nordgård, T. (ed.) *Nodalida '99 Proceedings from the 12<sup>th</sup> Nordiske datalingvistikkdager*, Department of Linguistics, University of Trondheim, p. 28-40.
- Hagen, K., J.B. Johannessen and P. Lane. 2001: Some problems related to the development of a grammar checker. Paper presented at *NODALIDA '01, the 2001 Nordic Conference in Computational Linguistics, May 21-22, 2001*.
- Hagen, K., J.B. Johannessen and A. Nøklestad. 2000a. The shortcomings of a tagger. In Nordgård, T (red.) *Nodalida '99 Proceedings from the 12<sup>th</sup> "Nordiske datalingvistikkdager"*, Department of Linguistics, University of Trondheim, p. 66-75.
- Hagen, K., J.B. Johannessen and A. Nøklestad. 2000b. A Constraint-based Tagger for Norwegian. I Lindberg, Carl-Erik and Steffen Nordahl Lund (red.): *17th Scandinavian Conference of Linguistics. Odense Working Papers in Language and Communication 19*, 31-48, University of Southern Denmark.
- Karlsson, F., A. Voutilainen, J. Heikkilä and A. Anttila. 1995. *Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin.
- Schneider, D. and K. McCoy. 1998. Recognizing Syntactic Errors in the Writing of Second Language Learners. In *COLING-ACL '98. Proceedings of the Conference*.

---

<sup>5</sup> This additional testing has been performed in cooperation with the project A Grammar Checker for Users with Special Needs, headed by Torbjørn Nordgård at NTNU.