

ADAPTING SIMPLENLG TO GERMAN

Marcel Bollmann
bollmann@linguistics.rub.de

Overview of the System

- Surface realisation engine for German, written in Java
- Based on version 3.8 of SimpleNLG for English (Gatt and Reiter, 2009) <http://code.google.com/p/simplenlg>
 - A port for the new architecture of version 4.x is planned
- Input specification:** lexical items, phrase specifications, grammatical features (e.g. verb tense, order of complements); alternatively: canned text
- Tasks performed by SimpleNLG for German:** ensuring correct orthography and grammatical well-formedness, including correct inflection, agreement, and word order (e.g. verb-second property)

Dealing with Word Order Variation

- Word order in German declarative sentences is restricted by verb-second property, but otherwise flexible:

Dem Mann gibt Maria ein Buch. 'Maria gives a book to the man.'
Ein Buch gibt Maria dem Mann.

- Word order should be controllable by the user
 - "Default" word order depends on semantic and syntactic factors and is therefore hard to always determine automatically
 - Variation in word order can be used for emphasis

- Argument realisation in SimpleNLG for English:
 - Subject at sentence level, complements at verb phrase level
 - Verb phrase realised as a single unit, i.e. movement of subject between two VP elements not possible without technical modifications
- Argument realisation in SimpleNLG for German:
 - All arguments, including the subject, realised in the verb phrase
 - Enables free ordering of arguments; also facilitates the implementation of passive constructions (e.g. raising objects to subject status)
 - Arguments for VP-internal subject position in, e.g., Haider (1993, p. 142 ff.)

Morphology

- Inflection is encapsulated in separate objects (*inflection patterns*) that largely resemble inflection classes from traditional grammars
- Inflection patterns consist of a set of suffixes that are appended to a stem and, optionally, a set of features

	NOM.	ACC.	DAT.	GEN.
MASC. SG.	—	-en	-em	-es
FEM. SG.	-e	-e	-er	-er
NEUT. SG.	—	—	-em	-es
PLURAL	-e	-e	-en	-er

An inflection pattern for possessive pronouns

- Features are used for, e.g., triggering plural umlaut or 'e' elision in word stems:

Buch 'book' → *Bücher* 'books'
sammeln 'to collect' → *ich sammle* 'I collect'

- New patterns can be defined by the user to enable, e.g., foreign inflectional paradigms in loanwords or (colloquial) inflectional variants, if desired

- Separable verbs can be instantiated by marking the boundary with a vertical bar; only the verb stem (following the bar) needs to be listed in the lexicon

`SPhraseSpec s2 = lex.makeS("Maria", "auf|heben", buch);`
→ *Maria hebt ein Buch auf.* 'Maria picks up a book.'

- Compound nouns can be instantiated in the same way; e.g. "Heimat|stadt"

Modal Verbs

- Verb phrase objects can take any number of modal verbs:
Sie sollte es sehen können müssen. 'She should be able to see it.'

- In modal constructions, the perfect tense setting affects the finite modal verb:

`s3.setPerfect(true);`
→ *Sie hat es tun können.* 'She was able to do it.'

- Alternatively, a separate feature can be used to set the main verb to perfect:

`s3.setMainVerbPerfect(true);`
→ *Sie kann es getan haben.* 'She might have done it.'

Basic Example

- As German is highly inflected, the use of a lexicon is mandatory; for testing and evaluation purposes, an adaption of IMSLex (Fitschen, 2004) has been used
- Assuming the variable *lex* has been initialized with a lexicon object, a basic sentence could be instantiated like this:

```
NPPhraseSpec mann = lex.makeNP("der", "Mann");
NPPhraseSpec buch = lex.makeNP("ein", "Buch");
SPhraseSpec s = lex.makeS("Maria", "geben", buch);
s.addIndirectObject(mann);
→ Maria gibt dem Mann ein Buch.                      'Maria gives a book to the man.'
```

- Two-layered system: "word order" for complements, "position" for modifiers
- "Word order" feature of the verb phrase controls the order of complements
 - Can be any permutation of subject (S), direct object (O), indirect object (I); genitive objects (which are relatively rare) are treated as direct objects
 - If not explicitly set, default word order is SIO

`s.setWordOrder(ISO);`
→ *Dem Mann gibt Maria ein Buch.* 'Maria gives a book to the man.'

- "Position" feature controls the placement of modifiers
 - Position of a modifier can be specified as either absolute (front, back) or relative to a complement (e.g., pre-subject or post-subject)
 - Multiple elements in the same position slot are realised in the order in which they were added to the sentence
 - Placement specification is obeyed even if complement word order changes

`s.addModifier(PRE_OBJECT, "heute");`
→ *Dem Mann gibt Maria heute ein Buch.* 'Today, Maria gives a book to the man.'

- The first constituent according to this ordering scheme is realised in the vorfeld position; all other constituents are realised in the mittelfeld

Relative Clauses

- Relative pronouns in German agree with their antecedent in gender and number, but inflect for case based on their function in the relative clause
- Utility functions have been added to facilitate relative clause creation; they allow specification of either:

- the grammatical function of the RP within the RC :
das Buch, das ich gekauft habe 'the book which I have bought'

- a preposition which should govern the RP
die Frau, auf die ich stolz bin 'the woman of whom I am proud'

- a noun phrase which the RP should specify, indicating possession.
die Frau, deren Kind schön ist 'the woman whose child is beautiful'

- Alternatively, sentence objects representing RCs can be constructed manually; this can be used to generate extraposed relative clauses

Evaluation and Future Work

- It is not obvious how a realisation engine can be systematically evaluated
- A quick, non-representative survey using Wikipedia articles suggested that 115 of 152 sentences (75.7%) were covered by the system's grammar

- Grammatical constructions which are not yet (fully) supported:
 - coordinated phrases (especially coordinated verb phrases)
 - verb cluster fronting

Gesehen hatte er sie nicht. 'He had not seen her.'

- semi-modal verbs requiring an infinitive with 'zu'

Er scheint sie zu kennen. 'He seems to know her.'

- Negation is currently confined to the insertion of the negation particle '*nicht*' at a fixed position; could possibly be refined

- Using canned text for proper nouns is problematic, as specifying gender/number information is often required for agreement purposes; this impedes the intended "simplicity of use" to a certain extent

Availability

- The full Java package of SimpleNLG for German will be made available online after it has been prepared for release