

Design and Implementation of a Surface Realiser for German based on the Architecture of SimpleNLG

Kira Klimt, 11.03.19, Master's Thesis Final Presentation

Advisor: Daniel Braun

Chair of Software Engineering for Business Information Systems (sebis)

Faculty of Informatics

Technische Universität München

www.matthes.in.tum.de

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

“Design and Implementation of a **Surface Realiser** for German based on the Architecture of SimpleNLG”

Natural Language Generation (NLG)

“NLG [...] is concerned with the construction of computer systems that can produce understandable text in English or other human languages from some underlying non-linguistic representation of information¹.”

NLG Process²



¹ E. Reiter and R. Dale. 1997. [Building Applied Natural Language Generation Systems](#).

² A. Gatt and E. Krahmer. 2018. [Survey of the state of the art in natural language generation](#).

Introduction

What is a Surface Realiser?

“Design and Implementation of a **Surface Realiser** for German based on the Architecture of SimpleNLG”

Input:

- *Subject*: Peter
- *Verb*: see
- *Object*: car
- *Determiner*: the
- *Tense*: past



Output:

Peter saw the car.

Introduction

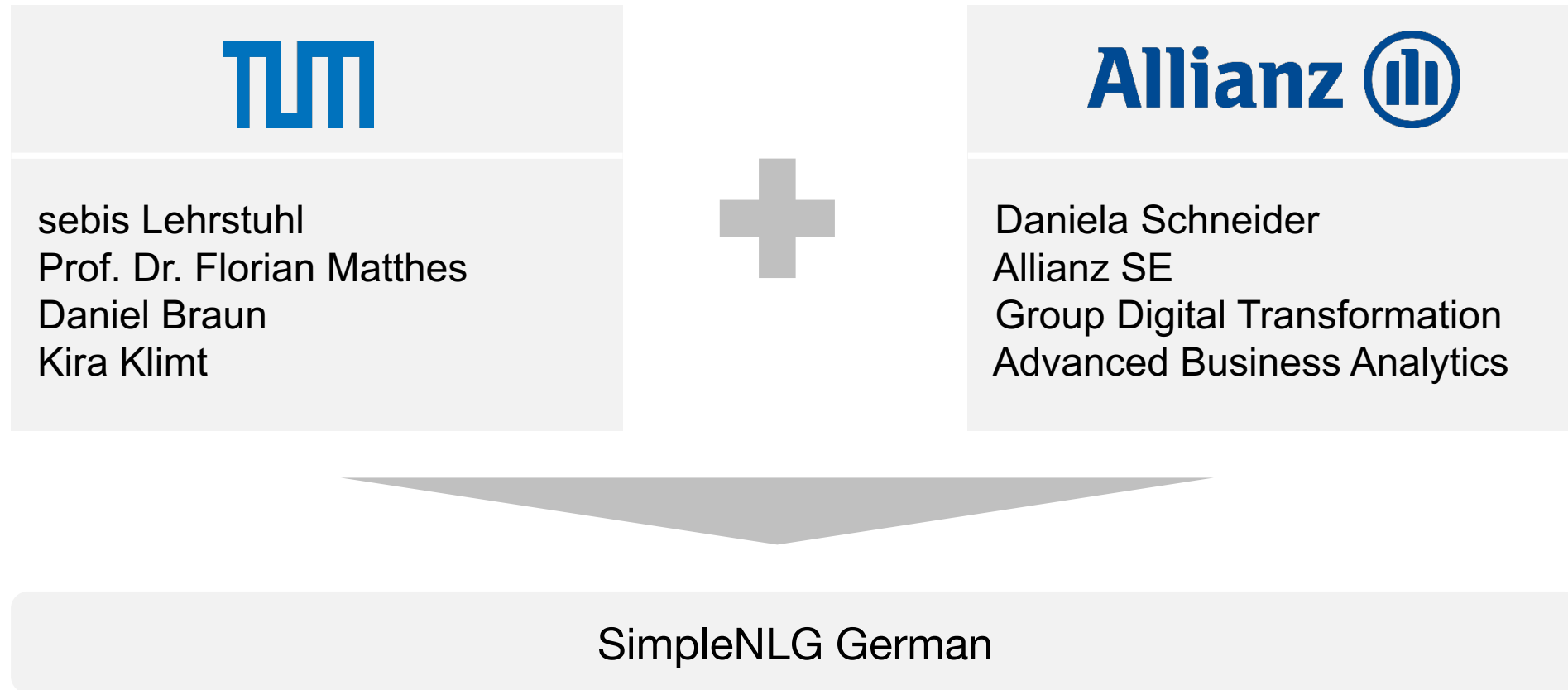
What is SimpleNLG?

“Design and Implementation of a Surface Realiser for German **based on the Architecture of SimpleNLG**”

SimpleNLG¹ is a realisation engine for the English language, which is:

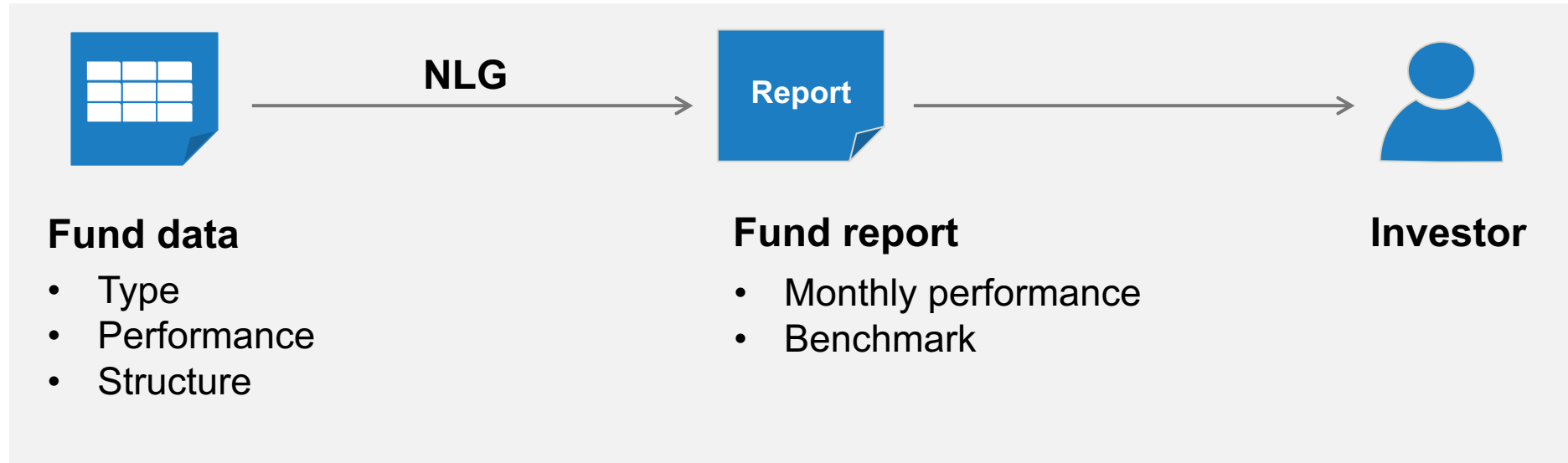
- A Java library
- Open Source
- Domain independent
- Simple and robust
- Adapted to 9 other languages
- Widely used

¹ A. Gatt and E. Reiter. 2009. [SimpleNLG: A realisation engine for practical applications.](#)



Introduction

Use case: Fund reporting



Current situation

- Much effort for manual reporting
- Template-based approach for share based funds (prototype)
- No approach for other types of funds

Why implementing a German realizer?

- Existing open source realisers (SimpleNLG) for English, French, Spanish,...
- Shortage of realisers for German
- German version of SimpleNLG existing, but incomplete, only minimally tested, based on outdated SimpleNLG version (not commercially utilizable)¹
- Comprehensive grammar rules and many special cases for German language
- Comprehensive Open Source Lexicon available

Why based on SimpleNLG? ²

- Open Source
- Not domain specific
- Offers wording and output control
- Simple usage
- Widely used

¹ M. Bollmann. 2011. [Adapting SimpleNLG to German](#).

² A. Gatt and E. Reiter. 2009. [SimpleNLG](#).

1 Introduction

2 Approach

3 Evaluation

4 Conclusion

5 Demo

1 Introduction

2 Approach

3 Evaluation

4 Conclusion

5 Demo

1

How should the **architecture** of a Surface Realiser look like?

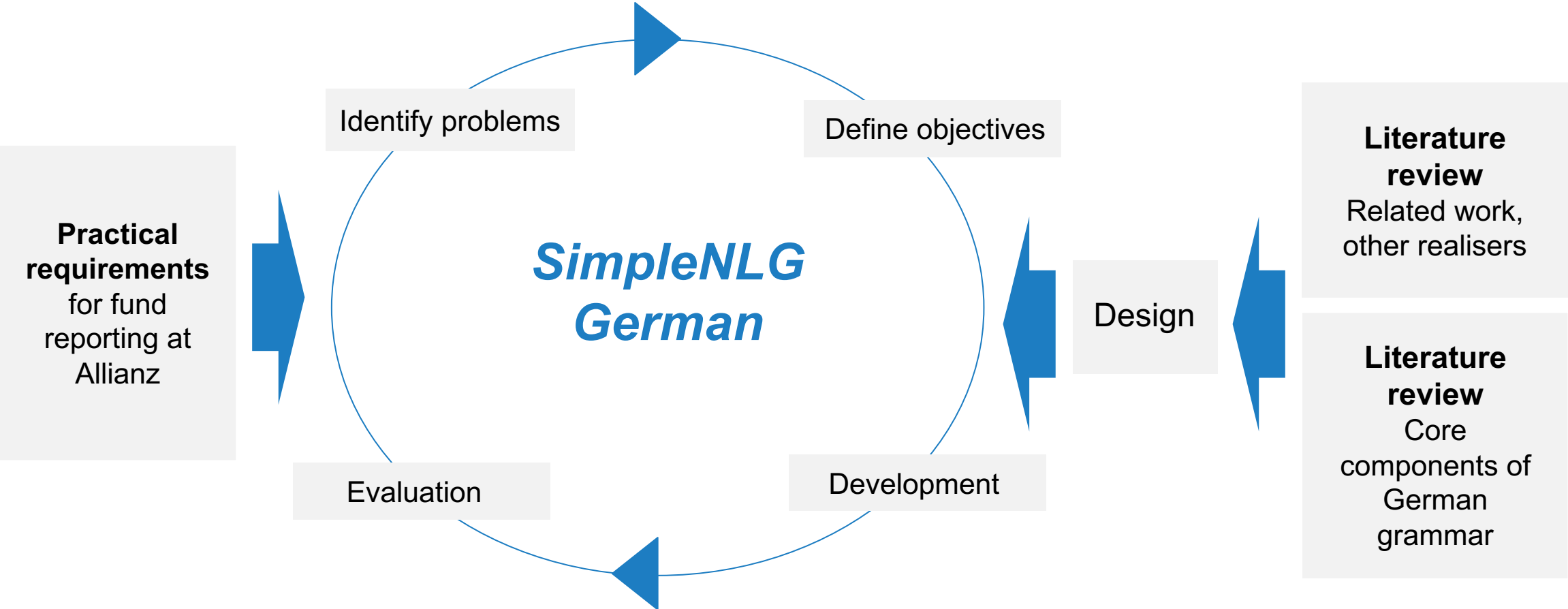
- What are **existing architectures**?
- Which role does **language** play with regard to the architecture?

2

Which **grammatical rule sets** need to be implemented in order to cover major parts of the German language?

3

How can we **evaluate** a Surface Realiser for the German language with regard to **quality and coverage**?



Business requirements

- Regular, irregular & separable verb conjugation
- Present & past tense
- Grammatical cases
 - Noun pluralization & declination
 - Adjective declination
- Enumerations
- Inflection of word group lexemes (e.g. Vereinigte Arabische Emirate)
- Simple usage
- Efficient realization
- Robustness

Requirements from literature review^{1,2}

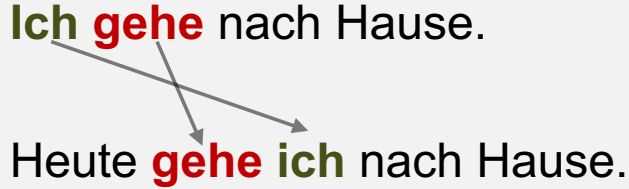
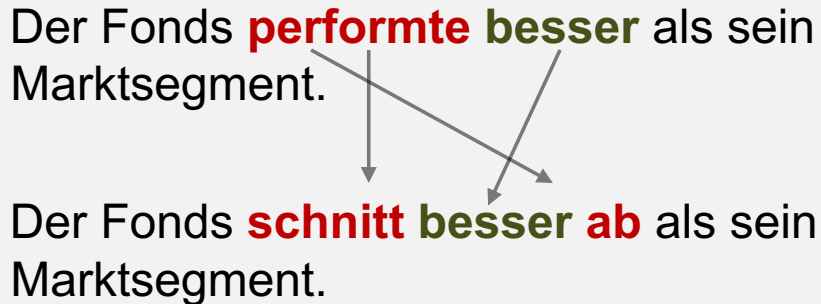
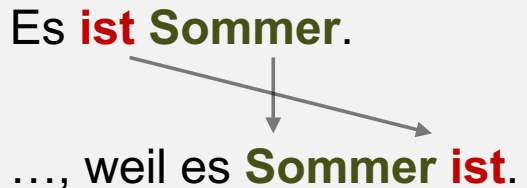
- Additionally, future and perfect tense
- Modal verbs
- Change of word order with front modifier
- Subordinate clauses
- Preposition contraction

¹ I. Balcik and K. Röhe. 2006. [PONS Deutsche Rechtschreibung und Grammatik](#).

² P. Eisenberg, J. Peters et al. 2006. [Duden - Die Grammatik](#).

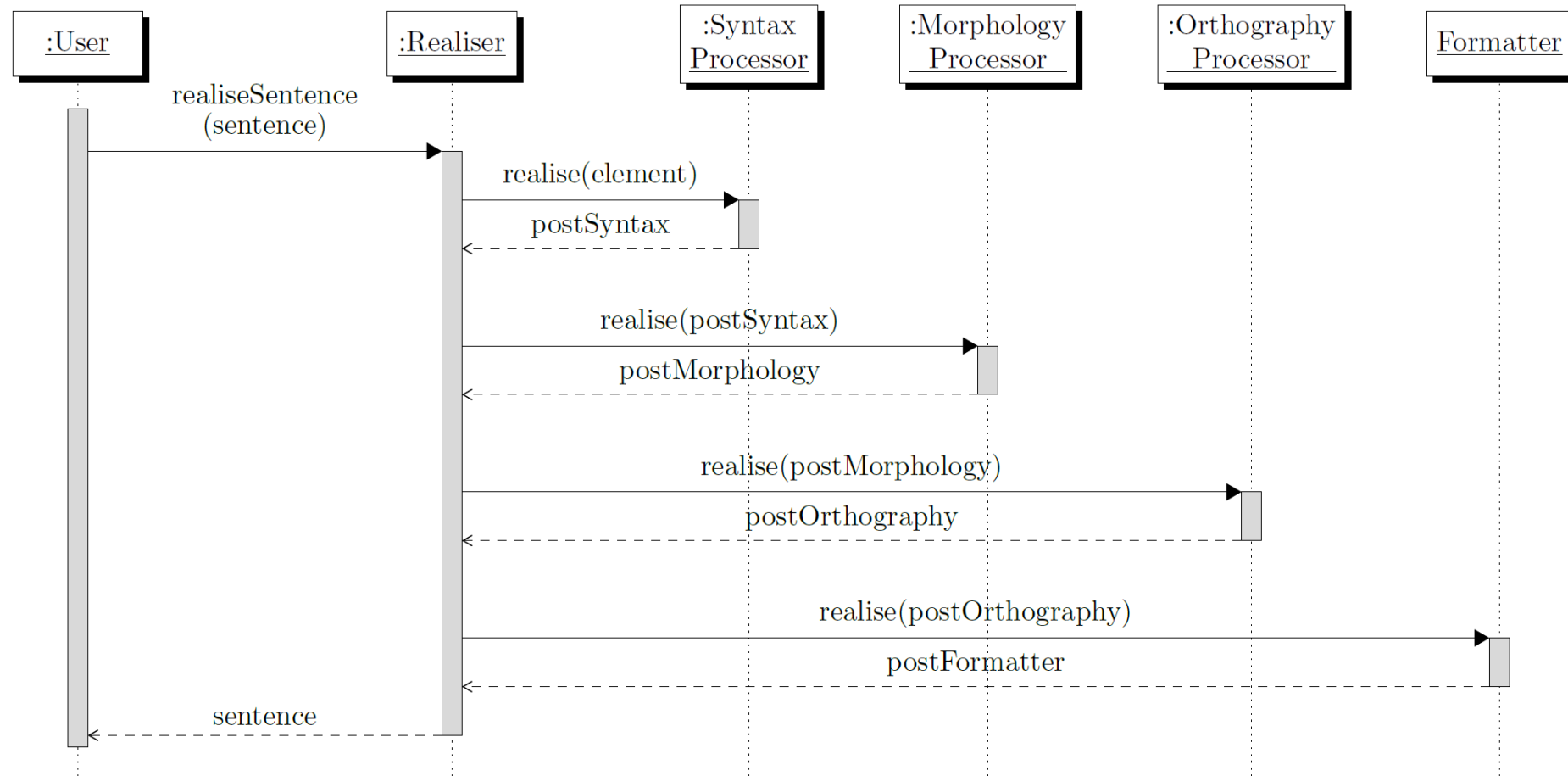
Approach

Why is German more complex than English?

<i>Issue</i>	<i>Input change</i>	<i>Example</i>
FrontModifiers change word order	 <code>setFrontModifier("Heute")</code> 	Ich gehe nach Hause. Heute gehe ich nach Hause. 
Separable verbs have to be split and change word order	 Change „performen“ to „abschneiden“ 	Der Fonds performte besser als sein Marktsegment. Der Fonds schnitt besser ab als sein Marktsegment. 
Initiated subordinate clauses change word order	 Add sentence as subordinate 	Es ist Sommer. ..., weil es Sommer ist. 

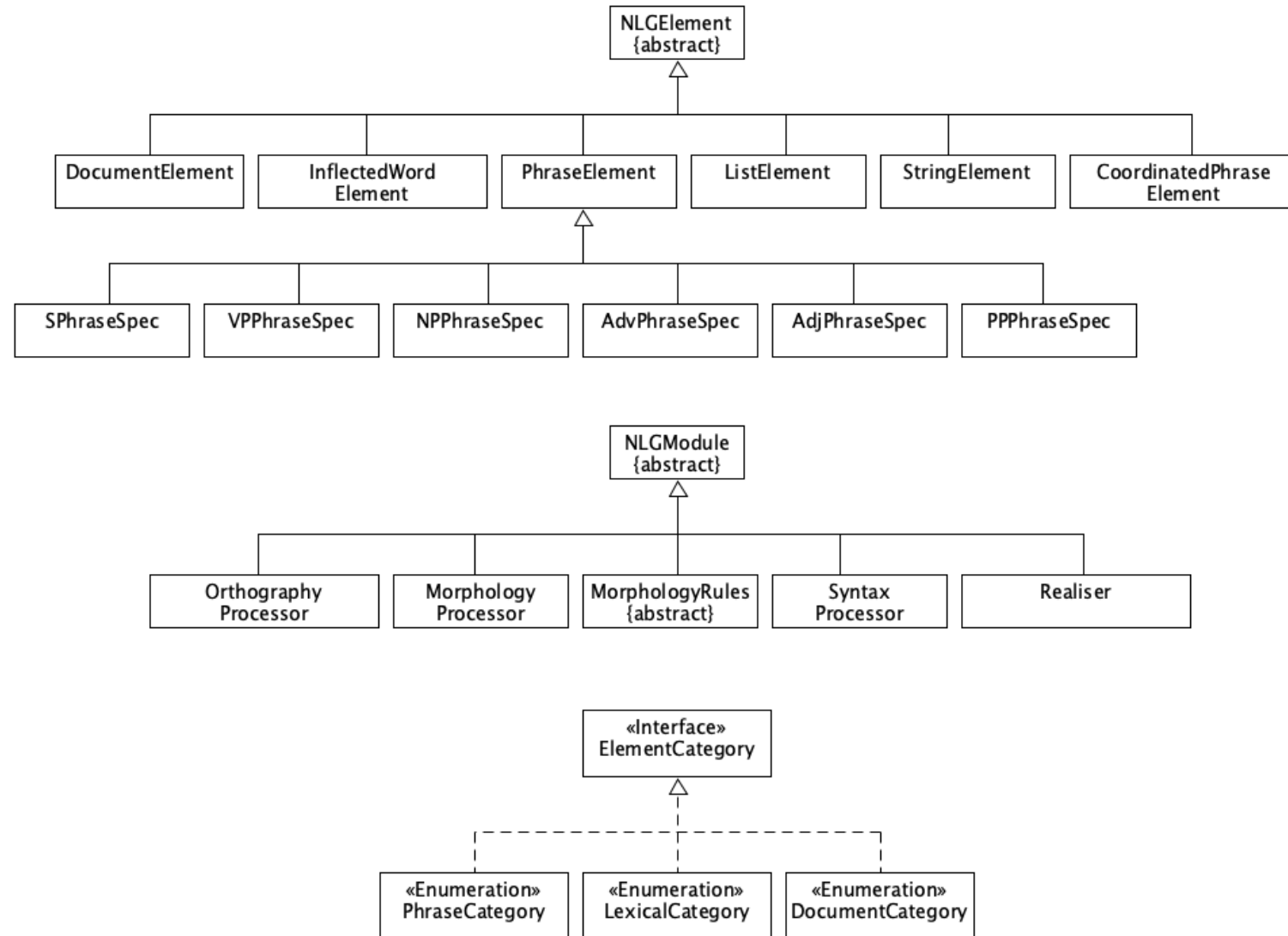
Approach

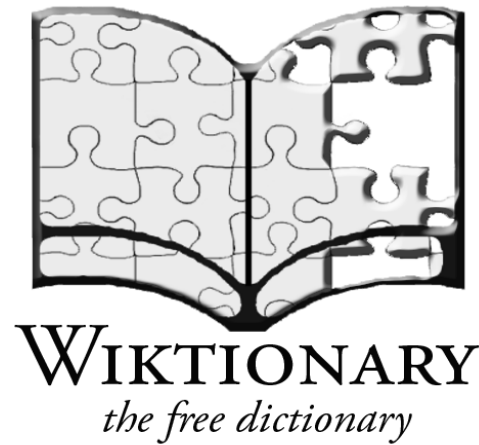
Architecture



Approach

Architecture





SimpleNLG German - *Lexicon* -

- Parsed 1,1 GB XML dump with 36.854.928 lines and large amounts of unneeded information
- To structured lexicon with 101.693 words
- Result: 79.866 nouns, 9803 verbs, 10.988 adjectives, 1036 adverbs
- Other German realisers: 125 nouns (OpenCCG)¹, 56 nouns (previous SimpleNLG German)²

¹ J. Vancoppenolle, E. Tabbert, G. Bouma, and M. Stede. 2011. [A German Grammar for Generation in OpenCCG](#).

² M. Bollmann. 2011. [Adapting SimpleNLG to German](#).

Wiktionary XML dump

```
<mediawiki xmlns="http://www.mediawiki.org/xml/export-0.10/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.mediawiki.org/xml/export-0.10.xsd" version="0.10" xml:lang="de">
  <page>
    <title>Flexion:widerspiegeln</title>
    <ns>0</ns>
    <id>1141</id>
    <revision>
      <id>6383478</id>
      <parentid>6316408</parentid>
      <timestamp>2018-06-04T12:07:20Z</timestamp>
      <contributor>
        <ip>194.44.179.182</ip>
      </contributor>
      <model>wikitext</model>
      <format>text/x-wiki</format>
      <text xml:space="preserve">== widerspiegeln (Konjugation) {{{Verbkonjugation|Deutsch}}} ==
{{{Deutsch Verb regelmäßig|Infinitiv=einteilig|Nebensatzkonjugation=einteilig|Teil 1=wider|spie|ge|l|n|widergespiegelt|vp|=|zp|=|gerund=|veraltet=0}}}
== widerspiegeln (Konjugation), reflexiv {{{Verbkonjugation|Deutsch}}} ==
{{{Deutsch Verb schwach trennbar reflexiv|wider|spie|ge|l|n|widergespiegelt|zr=ja}}}
</text>
      <sha1>2c6lcwnyqumiro7bgiojad8clozu57r</sha1>
    </revision>
  </page>
  <page>
    <title>widerspiegeln</title>
    <ns>0</ns>
    <id>1141</id>
    <revision>
      <id>6383478</id>
      <parentid>6316408</parentid>
      <timestamp>2018-06-04T12:07:20Z</timestamp>
      <contributor>
        <ip>194.44.179.182</ip>
      </contributor>
      <model>wikitext</model>
      <format>text/x-wiki</format>
      <text xml:space="preserve">== widerspiegeln {{{Sprache|Deutsch}}} ==
== {{Wortart|Verb|Deutsch}} ==

{{{Deutsch Verb Übersicht
|Präsens_ich=spiegle wider
|Präsens_ich*=spiegel wider
|Präsens_ich*=spiegele wider
|Präsens_du=spiegelst wider
|Präsens_er, sie, es=spiegelt wider
|Präteritum_ich=spiegelte wider
|Partizip_I=widergespiegelt
|Konjunktiv II_ich=spiegelte wider
|Imperativ Singular=spiegle wider
|Imperativ Singular*=spiegel wider
|Imperativ Singular*=spiegele wider
|Imperativ Plural=spiegelt wider
|Hilfsverb=haben
}}}

{{{Worttrennung}}}
:Wi·der·spie·geln, {{Prät.}} spie·gel·te wi·der, wi·der·spie·gel·te, {{Part.}} wi·der·ge·spie·gelt

{{{Aussprache}}}
:{{{IPA}}} {{{Lautschrift|ˈviːdə ˈʃpiːɡlən}}}
:{{{Hörbeispiele}}} {{{Audio|De-widerspiegeln.ogg}}}

{{{Bedeutungen}}}
:{{1}} {{{K|trans.}}} von etwas ein [[Spiegelbild]] [[werfen]], etwas reflektieren
:{{2}} {{{K|refl.}}} von etwas als Spiegelbild gezeigt werden, reflektiert werden
:{{3}} {{{K|trans.}}|übertr.}} etwas (oft indirekt) [[darstellen]], [[erkennen]] [[lassen]], von etwas [[prägen|geprägt]] sein
:{{4}} {{{K|refl.}}|übertr.}} in etwas (indirekt) zum Ausdruck kommen, erkennbar werden, etwas prägen

{{{Herkunft}}}
:{{Derivation}} des Verbs ''[[spiegeln]]'' mit dem Präfix ''[[wider-]]''

{{{Synonyme}}}
:{{1}} [[reflektieren]]. [[spiegeln]]. [[widerscheinen]]. [[widerstrahlen]]. [[zurückwerfen]]
```

SimpleNLG lexicon

```
<word>
  <base>widerspiegeln</base>
  <id>9</id>
  <category>verb</category>
  <regular>True</regular>
  <separable>True</separable>
  <reflexive>True</reflexive>
  <part1>wider</part1>
  <preterite>spiegelte wider</preterite>
  <participle2>widergespiegelt</participle2>
  <firstPerPres>spiegele wider</firstPerPres>
  <secPerPres>spiegelst wider</secPerPres>
  <thirdPerPres>spiegelt wider</thirdPerPres>
</word>
<word>
  <base>Fonds</base>
  <id>10</id>
  <category>noun</category>
  <plural>Fonds</plural>
  <genus>m</genus>
  <genitive_sin>Fonds</genitive_sin>
  <genitive_pl>Fonds</genitive_pl>
  <dative_sin>Fonds</dative_sin>
  <dative_pl>Fonds</dative_pl>
  <akkusative_sin>Fonds</akkusative_sin>
  <akkusative_pl>Fonds</akkusative_pl>
</word>
<word>
```

Verb conjugation according to person & tense - Regular, irregular & separable verbs - Present, past, perfect, future - Passive present & past - Modal verbs in present	“er geht”, “er ist”, “er fährt ab” “sie ging”, “sie hat gemacht”, “sie wird machen” “es wird gemacht”, “es wurde gemacht” “wir wollen gehen”
Adjective declination according to case, number & article, regular & irregular	”hoch”: “das hohe Haus”, “dem hohen Haus”, “der hohen Mauer”, “der hohen Mauern”
Adjective & adverb comparative & superlative	”gut”, “besser”, “besten” “der schnellere Zug” vs. “der Zug fährt schneller”
Noun declination according to case & number, regular & irregular	“das Haus”, “die Häuser”, “den Häusern”, “des Hauses”
Article declination according to case, number & gender, direct & indirect articles	“ das Haus”, “ die Häuser”, “ ein Haus”, “ eines Hauses”
Inflection of word group lexemes	“aus den Vereinigten Arabischen Emiraten”
Preposition contraction detection	„ im großen Haus“

Declarative main clauses	„Mary macht Hausaufgaben in der Küche.“ "Mary arbeitet Hausaufgaben in der Küche ab."
Compound sentences	„Der Hund bellt und die Katze miaut.“ „Der Hund bellt, die Katze miaut.“
Initiated subordinate clauses & conjunctions - Temporal - Causal - Conditional - Consecutive - Concessive - Modal - Comparative - Final - Adversative	"Die Sonne scheint, während es regnet." "Die Sonne scheint, weil es Sommer ist." "Die Sonne scheint, wenn du brav bist." "Die Sonne scheint, sodass alle ins Schwitzen kommen." "Die Sonne scheint, obwohl es regnet." "Die Sonne scheint, indem sie Kernfusion betreibt." "Die Sonne scheint heller, als die Lampe brennt." "Die Sonne scheint, damit wir nicht frieren." "Der Hund bellt, wohingegen die Katze miaut."
Appositions	"SAP, eine deutsche Firma,..."
Enumerations	„SAP und Bayer“ „SAP, Bayer und EON“

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

Morphology evaluation

- *Regular, irregular & separable verbs*¹:
 - 50 verbs in total fully conjugated & tested
- *Adjectives*: 15 adjectives¹, in different article combinations tested.
 - Only class not realized correctly: foreign words like “makaber”
- *Nouns*: 10 irregular, 10 regular, 10 word group lexemes¹ tested. Problems:
 - Word group lexemes with unknown adjectives
 - Unknown compound nouns not in lexicon

In total: **3828 test phrases**.

Syntax evaluation

- *Main clauses* (subject, predicate, direct & indirect object) *with & without separable verb*:
 - “Mary macht Hausaufgaben in der Küche.”
 - “Mary arbeitet Hausaufgaben in der Küche ab.”
- *9 types of initiated subordinate clauses*:
 - “Der Hund bellt, wohingegen die Katze miaut.”
 - “Die Sonne scheint heller, als die Lampe brennt.”

Evaluation of the system as whole

- Semi-automatically generating SimpleNLG code from TIGER corpus annotations²
- 100 sentences: 84% grammatical correctness

¹ Wiktionary. Accessed 2019-02-27. [Verzeichnis:Deutsch](#).

² W. Lezius, P. Eisenberg et al. 2005. [TIGER: Linguistic Interpretation of a German Corpus](#).

Testing the system as a whole with TIGER corpus

TIGER corpus

```
<s id="s4">
  <graph root="s4_VROOT">
    <terminals>
      <t id="s4_1" word="Der" lemma="der" pos="ART" morph="Nom.Sg.Masc" case="Nom" number="Sg" gender="Masc" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_2" word="texanische" lemma="texanisch" pos="ADJA" morph="Pos.Nom.Sg.Masc" case="Nom" number="Sg" gender="Masc" person="--" degree="Pos" tense="--" mood="--" />
      <t id="s4_3" word="Milliardär" lemma="Milliardär" pos="NN" morph="Nom.Sg.Masc" case="Nom" number="Sg" gender="Masc" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_4" word="Ross" lemma="Ross" pos="NE" morph="Nom.Sg.Masc" case="Nom" number="Sg" gender="Masc" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_5" word="Perot" lemma="Perot" pos="NE" morph="Nom.Sg.Masc" case="Nom" number="Sg" gender="Masc" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_6" word="hat" lemma="haben" pos="VAFIN" morph="3.Sg.Pres.Ind" case="--" number="Sg" gender="--" person="3" degree="--" tense="Pres" mood="Ind" />
      <t id="s4_7" word="das" lemma="der" pos="ART" morph="Acc.Sg.Neut" case="Acc" number="Sg" gender="Neut" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_8" word="politische" lemma="politisch" pos="ADJA" morph="Pos.Acc.Sg.Neut" case="Acc" number="Sg" gender="Neut" person="--" degree="Pos" tense="--" mood="--" />
      <t id="s4_9" word="Establishment" lemma="Establishment" pos="NN" morph="Acc.Sg.Neut" case="Acc" number="Sg" gender="Neut" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_10" word="in" lemma="in" pos="APPR" morph="--" case="--" number="--" gender="--" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_11" word="Washington" lemma="Washington" pos="NE" morph="Dat.Sg.Neut" case="Dat" number="Sg" gender="Neut" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_12" word="aufgeschreckt" lemma="aufschrecken" pos="VVPP" morph="Psp" case="--" number="--" gender="--" person="--" degree="--" tense="--" mood="--" />
      <t id="s4_13" word="." lemma="--" pos="$. " morph="--" case="--" number="--" gender="--" person="--" degree="--" tense="--" mood="--" />
    </terminals>
  </graph>
</s>
```

SimpleNLG code

```
53 SPhraseSpec sentence2 = nlgFactory.createClause();
54 NPPhraseSpec word11 = nlgFactory.createNounPhrase("der Milliardär");
55 word11.addModifier("texanisch");
56 NPPhraseSpec word12 = nlgFactory.createNounPhrase("Ross Perot");
57 word11.addPostModifier(word12);
58
59 VPPhraseSpec word14 = nlgFactory.createVerbPhrase("aufschrecken");
60 word14.setFeature(Feature.PROGRESSIVE, true);
61
62 NPPhraseSpec word15 = nlgFactory.createNounPhrase("das Establishment");
63 word15.addModifier("politisch");
64 word15.addPostModifier("in");
65 NPPhraseSpec word16 = nlgFactory.createNounPhrase("Washington");
66 word15.addPostModifier(word16);
67
68 sentence2.setFeature(Feature.TENSE, Tense.PERFECT);
69 sentence2.setSubject(word11);
70 sentence2.setVerb(word14);
71 sentence2.setObject(word15);
72 String output2 = realiser.realiseSentence(sentence2);
```

- 1 Introduction
- 2 Approach
- 3 Evaluation
- 4 Conclusion
- 5 Demo

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

What are the key findings? How to continue?



Key findings

- Splitting realisation process into syntax, morphology & orthography works well for German language
- Comprehensive lexicon important for German
- Separable verb splitting essential
- Comprehensive testing on different types of text, genres and feature combinations essential



Future Work

- Compound word splitting
- Splitting of unknown separable verbs by prefix detection
- Separable verbs in subordinate clauses
- Interrogative & imperative sentences
- Support diverse word orders & different placement of subordinate clauses
- Further tenses (future II, pluperfect, modal verbs & passive in further tenses)
- Further test automation

- I. Balcik and K. Röhe, PONS Deutsche Rechtschreibung und Grammatik, 1st ed. Stuttgart: Ernst Klett Sprachen GmbH, 2006.
- M. Bollmann, “Adapting SimpleNLG to German,” in Proceedings of the 13th European Workshop on Natural Language Generation (ENLG), 2011, pp. 133–138.
- P. Eisenberg, J. Peters, P. Gallmann, C. Fabricius-Hansen, D. Nübling, I. Barz, T. A. Fritz, and R. Fiehler, Duden - Die Grammatik. Unentbehrlich für richtiges Deutsch, 7th ed., K. Kunkel-Razum and F. Münzberg, Eds. Mannheim: Dudenverlag, 2006.
- A. Gatt and E. Krahmer, “Survey of the state of the art in natural language generation: Core tasks, applications and evaluation,” Journal of Artificial Intelligence Research, vol. 61, no. c, pp. 1–64, 2018.
- A. Gatt and E. Reiter, “SimpleNLG : A realisation engine for practical applications,” in Proceedings of the 12th European Workshop on Natural Language Generation, March, 2009, pp. 90–93.
- W. Lezius, P. Eisenberg, G. Smith, S. Brants, E. König, C. Rohrer, S. Hansen-Schirra, H. Uszkoreit, and S. Dipper, “TIGER: Linguistic Interpretation of a German Corpus,” Research on Language and Computation, vol. 2, no. 4, pp. 597–620, 2005.
- E. Reiter and R. Dale, “Building applied natural language generation systems,” Natural Language Engineering, vol. 3, no. 1, pp. 57–87.
- J. Vancoppenolle, E. Tabbert, G. Bouma, and M. Stede, “A German Grammar for Generation in OpenCCG,” in Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL), Hedeland Hanna, Schmidt Thomas, and Wörner Kai, Eds., 2011, pp. 145–150.
- Wiktionary German, “Verzeichnis:Deutsch,” Accessed 2019-02-27. [Online]. Available: <https://de.wiktionary.org/wiki/Verzeichnis:Deutsch>

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo

1

Introduction

2

Approach

3

Evaluation

4

Conclusion

5

Demo



B.Sc.

Kira Klimt

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Kira.klimt@tum.de



BACKUP

Adjectives with foreign origin with e-deletion

- *makaber* → *eine makabre Geschichte*, not *eine makabere Geschichte*

Compound nouns with irregular inflection not in the lexicon

- *Handelsverband* → *Handelsverbände*, not *Handelsverbanden*

Word group lexemes with adjectives not in the lexicon

- *Passauer Land* → *aus dem Passauer Land*, not *aus dem Passauren Land*

Correct sentences with different word order

Sentence after realisation

- *Und auch **halten** die Konzernchefs in den USA nicht viel von dem 62jährigen.*
- *Doch **geht** der Fortschritt dem Regierungschef nicht schnell genug.*



Sentence in corpus

- *Und auch die Konzernchefs in den USA **halten** nicht viel von dem 62jährigen.*
- *Doch der Fortschritt **geht** dem Regierungschef nicht schnell genug.*

Sentence after realisation

- *Und **hineinreichen** sie in das Unverständnis über den Problemkomplex multikulturelle Gesellschaft.*
- *Da **hermüssen** Schätzungen.*
- *Ungefähr die Hälfte des Schadens soll durch **Betriebsfremden** verursacht sein.*
- *Waitzis rät stattdessen zum lauten Protest in den Geschäften selbst und zu Beschwerden bei den **Handelsverbänden**.*
- *Gegen gewaltbereite und rechtsradikale Minderheiten wird nur der Staatsanwalt **hilft kann**.*

Problem

Separable verb not in lexicon

Compound noun not in lexicon & irregular plural form

Modal verbs in future tense currently not supported

DocumentElement: *category=SENTENCE realisation=Der Betreiber gewährt vierzehn Tage Widerrufs*

| -ListElement: *features={components=[der, Betreiber] case=SUBJECT }*

| | -NLGElement: *der*

| | \-NLGElement: *Betreiber*

| \-ListElement: *features={separable=false components=[gewährt, [[[vierzehn], Tage]], Widerrufsrecht]] discourse_function=VERB_PHRASE }*

| | -NLGElement: *gewährt*

| | \-ListElement: *features={components=[[[[vierzehn], Tage]], Widerrufsrecht] discourse_function=COMPLEMENT case=OBJECT }*

| | | -ListElement: *features={components=[[vierzehn], Tage] }*

| | | | -ListElement: *features={components=[[vierzehn], Tage] discourse_function=FRONT_MODIFIER }*

| | | | | -ListElement: *features={components=[vierzehn] }*

| | | | | | -NLGElement: *vierzehn*

| | | | | \-NLGElement: *Tage*

| | | \-NLGElement: *Widerrufsrecht*