Deriving Selectional Preferences And Verb Valency From Syntactically Annotated Corpora

Hybrid Machine Translation
SoSe 2018
Aleksandra Piadina
Jana Murasová
Table of Content

1. Valency (Verb Valency) Definition
2. Monolingual Valency Lexicography
3. Bilingual Valency Lexicography
4. Valency Extraction From Corpora
5. Valency in Machine Translation
6. CroCo-Corpus
7. Parallel Valency Derivation
8. Empty Links Phenomenon
9. Crossing Lines Phenomenon
Valency (verb valency)

• Verbs determine the sentence structure.

• Valency
  • Semantic level: arguments controlled by a verbal predicate (subject, direct/indirect object, prepositional object).
  • Syntactic level: syntactic realization of arguments.
Monolingual Valency Lexicography

• German Valency Dictionary by Helbig und Schenkel (1969)
  • first German valence dictionary
  • for non-native German speaker
  • 3-level information for a verb:
    • number of obligatory and optional arguments
      • danken1(2,3) (to thank)
    • morpho-syntactic properties of the arguments
      • danken → Sn,(Sd),(pS|NSdaß)
    • semantic types of the arguments
      • Sn → 1. Hum (der Jubilar dankt/The jubilarian thanks),
        2. Abstr (Der Betrieb dankt dem Ministerium/The company thanks the Ministry)
Monolingual Valency Lexicography

• **German Valency Dictionary by Engel und Schumacher (1976)**
  - only syntactic information
    - different types of arguments (Ergänzungsklassen) from E0 to E9 for a verb
      - essen 0(1 (to eat))
    - verb ability to passivization
    - SE argument types for arguments in sentence form (satzförmige Ergänzungen)
      - danken 013 (to thank)
        1:SE DASS „Niemand dankt es ihr, daß sie ihm geholfen hat.“
        („Nobody thanks her for helping him“)

• **English Valency Dictionary by Emons (1978)**
  - E. g.: believe
    \[S12[P12 + E1[NOM1/ES1] + [E2[NOM2/ES2[that]]]]\]
Monolingual Valency Lexicography

- **English Valency Dictionary by Allerton (1982)**
  - distinguishes between SUBJECT, OBJECT and OBJOID (objects which can’t be subjects in passivization, e.g. *Thomas is a student*)
  - E.g.: *see*
    - *Valency class 12*
    - SUBJECT + V + OBJECT
  - classification of argument types, e.g.
    - *reside*
      - *Thomas resided in the palace*
Monolingual Valency Lexicography

• **German Verb classification by Ballmer and Brennenstuhl (1986)**
  • motivation: classify German verbs (about 8000) according to a temporal causal model created by the authors
  • valency information for each class of verbs
  • verbs are grouped not only into classes but in models describing a process and representing different phases of the process
Monolingual Valency Lexicography

example of a model

- **Zustoßmodell (Happening)**
  - **Vorspielphase:** Sich einem Einfluss aussetzen jed1 („sich sonnen“)
    - initial situation: to expose yourself to an influence sb1 ("to sunbathe")
  - **Ablaufphase:** Zustoßen etw1 jd3 (allgemein: „geschehen, passieren“, speziell: „verbrennen“)
    - transition from initial to end situation: to happen smt1 to sb3 (in general: “to happen”, specifically: “to burn”)

- **Schlussphase:** Sich auswirken etw1 auf jd2 („aufregen, erschrecken“)
  - end situation: to have an effect smt1 on sb2 („to annoy, to scare“)

- **Nachspielphase:** Reagieren jd1 auf etw2 („antworten, reagieren“)
  - consequence: to react sb1 to smt2 („to answer, to react“)
Monolingual Valency Lexicography

• FrameNet/SALSA (English/German)
  • sentence annotation based on frame semantics
Monolingual Valency Lexicography

  - E. g.: see
    - ‘look’ Active 1/3 Passive 1/3 General 0
    - I [N]_{A\|by N}
    - II [N]_{P\|N\|P-only [thatCL]_{P(it)} [whCL]_{P(it)}}
    - III [for REFL PRON]
Bilingual Valency Lexicography

- German-Romanian Dictionary (Engel und Savin 1983)
  - semantic restrictions are necessary, e.g.
    - Schwimmen (to swim)
      - a înota for „living beings actively swimming“
      - a pluti for „floating objects in the water“

+ANIM for living beings (a înota +ANIM)
+MAT for inanimate beings (a pluti +MAT)
Bilingual Valency Lexicography

- **German-Portuguese Valency Dictionary by Welker (2003)**
  - motivation: represent valency entries in the way that even users with no specific linguistic knowledge can easily use it
  - abbreviations for the representation of arguments in valency frames which combine semantic and syntactic properties
    - P for persons, A for things, AN for an animal
    - Example for an entry: $P * P$
    - Subject
    - Object (Akkusativobject)
    - Verb

  (e.g. *lieben* – someone loves someone)
Bilingual Valency Lexicography

  - uses frame semantics annotations
  - E. g. walk
    - Bernd walked to the door. (Bernd ging zur Tür.)
      \[
      \text{walk} [\text{Self-mover}_{\text{NP}} \text{Goal}_{\text{NP}}] \quad \rightarrow \quad \text{gehen}
      \]
    - Bernd walked Anna to the door. (Bernd begleitete Anna zur Tür.)
      \[
      \text{walk} [\text{Self-mover}_{\text{NP}} \text{Cotheme}_{\text{NP}} \text{Goal}_{\text{PP}}] \quad \rightarrow \quad \text{begleiten}
      \]
"Selectional preferences characterize the potential arguments of word senses in terms of their semantic properties."

E. g.: the verb *eat* strongly prefers an object in the category of *food*
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Motivation:
  • handling ambiguity
    • pronoun resolution
    • word sense disambiguation
  • parsing (more “natural” arguments for words)
  • infer semantic properties of words missing from the lexicon
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Resources and Methods
  • The Rational Components:
    • The Parser
      • English Slot Grammar
        → complement slots (slot frame)
        • subj (subject)
        • obj (direct object)
        • iobj (indirect object)
        • comp (complement of object or subject)
        • auxcomp (auxiliary complement)
        • pred (predicate complement of “be”)
        → adjunct slots
ESG parse of “The Russian emperors ate caviar and drank vodka”
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• The Lexicon and Semantic Types
  • lexicon (94,000 base forms) with word senses marked with semantic types (approximately 450 types)
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• The Empirical Components:

  • **Robust Corpus Processing**
    • a corpus of unannotated Reuters newswire comprising approximately 6.4 million sentences

  • **Frequency and Maximum-Likelihood Estimation**
    • A variation on simple relative frequency determines the selectional preferences for complements.
    • Let the frequency of a specific slot frame \( f \) for a verb \( v \) in the training corpus be \( \text{freq}(f) \). The following then describes the simple relative frequency of a specific slot frame \( f_0 \):

\[
\frac{\text{freq}(f_0)}{\sum_{f \in F} \text{freq}(f)}
\]

where \( F \) is the set of frames for \( v \).
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Advantages of the Hybrid Approach
  • consideres full slot frames
    • E. g.: The cow ate it.
      where possible antecedents are mouse and grass
    Solution: store subject-verb-object frames

Sample slot frame output for the verb eat
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

- uses **class-based models**
- assign **probability values to classes of words** rather than to individual words
- semantic type → super semantic type T
- (e.g. *Artist* is replaced by *Human*)
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Experiments and Results
  • analyzed 6.4 million sentences from a corpus of Reuters newswire
    slot frames for 6760 verbs
  • comparison with the list of Resnik („Selectional constraints: An
    information-theoretic model and its computational realization .”1996)
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Approach of Resnik (1996)
  • „Selectional constraints are limitations on the applicability of natural language predicates to arguments.“
    ➢ specify what is or what is not an appropriate argument for a particular predicate
  • information-theoretic model of selectional constraints
    • generic taxonomic representation of concepts
e. g. wine – beverage – liquid
    • probabilistic formalization of selectional constraints computed on the basis of simple frequencies of co-occurrence between predicates and their arguments
    (probabilistic relationship between predicates and conceptual categories or classes)
• experiment testing the model’s ability to distinguish plausible from implausible direct objects of verbs
  • The Brown corpus was used as learning sample.
  • For each pairing of a verb $v$ and direct object $n$ (a set of 16 verbs having a bias for NP complements), the selectional association $A(v, c)$ was calculated for each WordNet class $c$ to which $n$ belongs.
  • The greatest such value of selectional association was assigned as the model’s rating for the pairing $(v, n)$. 
### Comparison

<table>
<thead>
<tr>
<th>Verb</th>
<th>Assoc object</th>
<th>WN Class object</th>
<th>SG Class object</th>
<th>SG Class subject</th>
<th>SelPref</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>see</td>
<td>5.79</td>
<td>&lt;entity&gt;</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.468</td>
<td>22 (47)</td>
</tr>
<tr>
<td>read</td>
<td>6.80</td>
<td>&lt;writing&gt;</td>
<td>&lt;st_document&gt;</td>
<td>&lt;human&gt;</td>
<td>0.624</td>
<td>73 (117)</td>
</tr>
<tr>
<td>hear</td>
<td>1.89</td>
<td>&lt;communication&gt;</td>
<td>&lt;speech_act&gt;</td>
<td>&lt;human&gt;</td>
<td>0.212</td>
<td>41 (193)</td>
</tr>
<tr>
<td>write</td>
<td>7.26</td>
<td>&lt;writing&gt;</td>
<td>&lt;st_document&gt;</td>
<td>&lt;human&gt;</td>
<td>0.950</td>
<td>132 (139)</td>
</tr>
<tr>
<td>urge</td>
<td>1.14</td>
<td>&lt;life_form&gt;</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.938</td>
<td>2340 (2496)</td>
</tr>
<tr>
<td>warn</td>
<td>4.73</td>
<td>&lt;person&gt;</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>1.000</td>
<td>79 (79)</td>
</tr>
<tr>
<td>judge</td>
<td>1.30</td>
<td>&lt;contest&gt;</td>
<td>&lt;st_interaction&gt;</td>
<td>&lt;human&gt;</td>
<td>0.524</td>
<td>11 (21)</td>
</tr>
<tr>
<td>teach</td>
<td>1.87</td>
<td>&lt;cognition&gt;</td>
<td>&lt;st_discipline&gt;</td>
<td>&lt;human&gt;</td>
<td>0.210</td>
<td>21 (60)</td>
</tr>
<tr>
<td>expect</td>
<td>0.59</td>
<td>&lt;act&gt;</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.366</td>
<td>26 (71)</td>
</tr>
<tr>
<td>repeat</td>
<td>1.23</td>
<td>&lt;communication&gt;</td>
<td>&lt;speech_act&gt;</td>
<td>&lt;human&gt;</td>
<td>0.582</td>
<td>32 (55)</td>
</tr>
<tr>
<td>understand</td>
<td>1.52</td>
<td>&lt;cognition&gt;</td>
<td>&lt;st_cognition&gt;</td>
<td>&lt;human&gt;</td>
<td>0.159</td>
<td>10 (63)</td>
</tr>
<tr>
<td>Not in Resnik’s list:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>measure</td>
<td>&lt;st_outcome&gt;</td>
<td>&lt;st_document&gt;</td>
<td>0.464</td>
<td>150 (323)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eat</td>
<td>&lt;st_food&gt;</td>
<td>&lt;human&gt;</td>
<td>0.746</td>
<td>135 (181)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>drink</td>
<td>&lt;st_liquid&gt;</td>
<td>&lt;human&gt;</td>
<td>0.882</td>
<td>60 (68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kill</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.869</td>
<td>2109 (2428)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;human&gt;</td>
<td>&lt;st_event&gt;</td>
<td>0.036</td>
<td>87 (2428)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;human&gt;</td>
<td>&lt;air_vehicle&gt;</td>
<td>0.028</td>
<td>68 (2428)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;human&gt;</td>
<td>&lt;st_weapon&gt;</td>
<td>0.012</td>
<td>29 (2428)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;human&gt;</td>
<td>&lt;st_animal&gt;</td>
<td>0.004</td>
<td>10 (2428)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>love</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.855</td>
<td>106 (124)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_place&gt;</td>
<td>&lt;human&gt;</td>
<td>0.145</td>
<td>18 (124)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>throw</td>
<td>&lt;st_event&gt;</td>
<td>&lt;human&gt;</td>
<td>0.576</td>
<td>19 (33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_artifact&gt;</td>
<td>&lt;human&gt;</td>
<td>0.424</td>
<td>14 (33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>describe</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.722</td>
<td>65 (90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_event&gt;</td>
<td>&lt;human&gt;</td>
<td>0.156</td>
<td>14 (90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_document&gt;</td>
<td>&lt;human&gt;</td>
<td>0.122</td>
<td>11 (90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>study</td>
<td>&lt;st_document&gt;</td>
<td>&lt;human&gt;</td>
<td>0.534</td>
<td>119 (223)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_cognition&gt;</td>
<td>&lt;human&gt;</td>
<td>0.224</td>
<td>50 (223)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_action&gt;</td>
<td>&lt;human&gt;</td>
<td>0.049</td>
<td>11 (223)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>attack</td>
<td>&lt;human&gt;</td>
<td>&lt;human&gt;</td>
<td>0.671</td>
<td>496 (739)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;st_place&gt;</td>
<td>&lt;human&gt;</td>
<td>0.099</td>
<td>73 (739)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

- Resnik’s list in comparison with Bernth and C. McCord
  - only relate an object to a verb
  - different semantic types
  - only one frame
A Hybrid Approach to Deriving Selectional Preferences
Arendse Bernth and Michael C. McCord

• Applying the Selectional Preferences
  • anaphora resolution
    • important by producing the right gender and semantic types for translations of pronouns
    • E. g.: The **food** was put on the table by the cook.
      He then sat down to **eat** it.

      Mary borrowed a new **book** from the library,
      but she has not found time to **read** it yet.

      The teacher couldn’t understand John’s **question** because it was too loud in the room. Then he had to **repeat** it.
Valency Extraction From Corpora

- Experiments with monolingual English texts, Briscoe (2001)
  - Extraction system:
    - HMMTagger,
    - lemmatizer
    - statistical parser
    - pattern extractor
    - pattern classifier
    - lexical filter
Valency Extraction From Corpora

• E. g.: *He attributed his failure, he said, to no-one buying his book.*

  → he
  → his failure
  → to no-one buying his book

• Problems:
  • (a) *He looked up the word.*
  (b) *He looked up the hill.*
Valency Extraction From Corpora

• Parallel valence frame extraction from Prague Czech-English Dependency Treebank (PCEDT)

• **PCEDT** is a part of the Penn Treebank, which has been extended by Czech translations, and where texts were annotated using the pattern of Prague Dependency Treebank.

→ 2 layers of dependency annotations:
- analytical (dependencies)
- tectogrammatical (captures linguistic meaning)
Valency Extraction From Corpora

• experiments by Bojar und Hajič (2005)
  • structural transfer
  • Parallel valency pairs are extracted from the dependency trees of PCEDT verb pairs, each with a list of modifications.
Valency Extraction From Corpora

• Collocations and Verbequivalents from EUROPARL by Duffner et. al. (2009)

  • E. g.: *einstellen*
    • *stop* and *cease*
    • possible collocations:
      • *stop*: paying, giving, funding; aid; executions, activities
      • *cease*: hostilities
Valency in MT

- **SUSY**
  - multilingual transfer-based system
  - uses syntactic valency information by the selection of the target language word and by the generation

- **METAL**
  - uses valency frames
  - statistical components by the choice of frames

- **EUROTRA**
  - Interface Structure (dependencies)
Parallel Valency Derivation

• Verb Valency Parallelism Assumption – the valency is parallel (across languages)

• EUROTRA: The predicate-argument structure transfer can cause various transformations, head-switching or omission of single arguments. However, predicate equivalence is assumed.
Parallel Valency Derivation

• To what extend in the syntactic level do the original text and the translation correlate?

• Which movements of functions and categories in the language pair English-German could we expect?
CroCo-Corpus

• Cross linguistic corpora (Neumann and Hansen-Schirra 2005)
• contains English and German originals and their translations

• English-German (E2G, EO and GTrans)
• German-English (G2E, GO and ETrans)

• automatic alignment: words
• manuell alignment: simple clauses and sentences
CroCo-Corpus - Register

8 text types (register):

- ESSAY – political essays on economics
- FICTION – fictional texts
- INSTR – instructional manuals
- POPSCI – popular scientific texts
- SHARE – corporate communication
- SPEECH – prepared speeches (political)
- TOU – tourism leaflets
- WEB - websites
CroCo-Corpus - Annotation

- **annotation:**
  - meta information following the TEI standard
  - PoS information using the TnT tagger with the STTS tag set for German and the Susanne tag set for English
  - morphology using MPRO (for both languages)
  - grammatical functions of the highest nodes in the sentence, manually annotated with MMAX2
Ideal case

• **alignment:**
  • source and target text
  • sentences
  • simple clauses
  • phrases
  • words

• not always => **empty links**
Empty Links

- elements without any equivalent in SL, resp. TL
Crossing Lines

- aligned elements with a counterpart embedded in higher units which are not aligned
Crossing Lines

for every sentencePair in sentencePairs
    slSentence := getSlSentence(sentencePair)
    tlSentence := getTlSentence(sentencePair)

    for every clause in getClauses(slSentence)
        alignedClause := getAlignedClause(clause)
        if (not isMember?(alignedClause, tlSentence))
            markCrossingLine()
        end
    end
end

# repeat the same for tlSentence

• pseudo-code for querying crossing lines
Valency extraction

1. identification of the semantic main verb
2. evaluation of surrounding words
3. derivation of verb complements, adjuncts and grammatical functions

pseudo-code for the extraction and querying of grammatical-functional valency patterns
Valency extraction

- # pairs function with a category
- ; splits SL-Verb and TL-Verb
- \(<=>\) determines the language (translation) direction

\[ 
\text{"leisten": subj\#np, dobj\#np, ; "make": subj\#np, dobj\#np, Finanzmärkte leisten einen nicht zu unterschätzenden Beitrag für das Wachstum einer Volkswirtschaft. Financial markets make an invaluable contribution to economic growth.} 
\]

\[ leisten(subj:np, dobj:np) \leftrightarrow make (subj:np, dobj:np) \]
Valency extraction

• no exact equivalent (*leisten x make*)
• valency carrier => noun-verb-combinations (*Beitrag leisten x make a contribution*)

"leisten": subj#np, dobj#np, ; "make": subj#np, dobj#np,

Financial markets make an invaluable contribution to economic growth.
Results: Empty Links

• sentence alignment: 99%

• single clauses alignment:
  • EO and ETrans contain more single clauses
  • EO and ETrans contain more empty links

• single clause alignment
Empty Links on the level of single clauses

- German adverbial connectors ⇔ subordinate clauses in English

a. [Deshalb machen hohe Abgaben Arbeit teuer] [und können doch nicht verhindern,] [dass unseren Sozialsystemen der Kollaps droht.] (GO_SPEECH_007)
b. [That is why] [high taxes make work expensive] [and yet cannot protect our social system from] [impending collapse.] (ETrans_SPEECH_007)
Empty Links on the level of single clauses

- embedded adjective phrases in German ⇔ relative or participle clauses in English

a. [Die Staats- und Regierungschefs der Europäischen Union haben in Göteborg erneut ihre Bereitschaft bekräftigt,] [die in Kyoto eingegangenen Verpflichtungen zur Verminderung der Treibhausgase zu erfüllen.] (GO_SPEECH_001)

b. [In Gothenburg the EU heads of state and government reaffirmed their willingness] [to fulfil the commitments] [they made in Kyoto] [to reduce greenhouse gases.] (ETrans_SPEECH_001)
Empty Links on the level of single clauses

- *zu-PP* in German $\Leftrightarrow$ *to*-single clauses in English

\begin{itemize}
  \item [Mittlerweile ist anerkannt,] [dass es zur Sicherung von Beschäftigung vor allem auf Flexibilität ankommt.] (GO_SPEECH_007)
  \item [It has now been recognized] [that flexibility is the most important factor] [when it comes] [to safeguarding jobs.] (ETrans_SPEECH_007)
\end{itemize}
Empty Links on the level of grammatical functions

- varying percentage of unmapped functions => depending on register
- EO-GTrans: more than 30% of English appositions and complements don’t have any equivalent in German
- GO-ETrans: German predicatcors and modal adverbs are rarely parallel translated in English
Empty Links on the level of grammatical functions

• English appositions ⇔ coordinate finite sentences in German

\[ a. \text{ Revenues rose 11\% to } $112\text{ billion, [a record]}_{\text{APPO.}} \text{ (EO\_SHARE\_004)} \]
\[ b. \text{ Der weltweite Umsatz stieg um 11\% auf } $112\text{ Mrd. und erreichte damit eine neue Rekordhöhe.} \text{ (GTrans\_SHARE\_004)} \]

• English specialized register tends to be recipient-oriented
• German specialized register tends to be content-oriented

(House 1997)
Empty Links on the level of grammatical functions

- subject complements in English $\leftrightarrow$ prepositional objects in German

\begin{itemize}
  \item \textit{Also for the second straight year, we were named [“The World's Most Respected Company”]}_{\text{COMPL}} by the Financial Times. (EO_SHARE_004)
  \item \textit{Ebenfalls zum zweiten Mal in Folge ernannte die Financial Times GE [zum “am meisten respektierten” Unternehmen der Welt]}_{\text{PROBJ}}. (GTrans_SHARE_004)
\end{itemize}

- verbs \textit{name, elect, make}
Empty Links on the level of grammatical functions

- subject complements in English $\leftrightarrow$ prepositional objects in German

\begin{align*}
a. \quad & \text{We are [pleased to present the 2001 Annual Report of} \\
& \quad \text{the American Institute for Contemporary German Studies} \ (AICGS) \left[ \text{COMPL.} (EO\_SHARE\_013) \right] \text{. (EO\_SHARE\_013)} \\
\text{b. \quad & Wir freuen uns, [Ihnen den Jahresbericht 2001 des} \\
& \quad \text{American Institute for Contemporary German Studies} \ (AICGS) \text{ präsentieren zu können} \left[ \text{PROBJ.} \right] \text{. (GTrans\_SHARE\_013)} \end{align*}
Empty Links on the level of grammatical functions

• finite verbs in English ⇔ non-finite verbs in German

a. *In that report, we [described]_{FIN} several challenges and opportunities that we felt were going to determine the agenda of German-American relations. (EO_SHARE_013)*

b. *In diesem Bericht werden verschiedene Herausforderungen und Gelegenheiten [beschrieben]_{PRED}, die unserer Meinung nach die Beziehungen der beiden Staaten bestimmen. (GTrans_SHARE_013)*

• f.e. **active** vs. **passive** clauses (recipient-oriented vs. content-oriented)
Empty Links on the level of grammatical functions

- finite verbs in English ⇔ non-finite verbs in German

\[ a. \quad We \text{ already } [\text{have}]_{\text{FIN}} \text{ that}! \quad (\text{EO\_SHARE\_004}) \]
\[ b. \quad Das \text{ alles } [\text{haben}]_{\text{FIN}} \text{ wir bereits } [\text{geschafft}]_{\text{PRED.}}. \quad (G\text{Trans\_SHARE\_004}) \]
Empty Links on the level of grammatical functions

- main verbs in English ⇔ modal adverbs in German

a. *We have continued our efforts to ease the suffering of families of lost colleagues.* (EO_SHARE_005)
b. *Unsere Anteilnahme und Hilfe gilt [weiterhin]_ADV_MOD den Familien, deren Angehörige dabei ihr Leben lassen mußten.* (GTrans_SHARE_007)
Crossing Lines between words and grammatical functions

- at least one word in the phrase is aligned to a word embedded in a higher unit with another grammatical function

- considered for:
  - adjectives
  - adverbs
  - nouns
  - verbs
Crossing Lines between words and grammatical functions

• direct objects embedded in a prepositional objects in German
  => direct objects embedded in a direct objects in English

a.  Er hat sich [darauf]_{PROBJ} verlassen, [dass wir von drin
    nen [sein Lächeln]_{EMBEDDED_DOBJ} sehen können]_{PROBJ}
    (GO_FICTION_007)

b.  He just assumed [we could see [his smile]_{EMBEDDED_DOBJ}
    from inside]_{DOBJ}. (ETrans_FICTION_007)
Crossing Lines between words and grammatical functions

• postmodification within a prepositional objects in German shifted to a premodification of a direct objects in English

\[ a. \quad 1995 \text{ haben wir } [\text{auf 125 Jahre Deutsche Bank}]_{\text{PROBJ zu rückgeblickt.}}. \ (\text{GO\_SHARE\_009}) \]
\[ b. \quad \text{In 1995 we celebrated } [\text{Deutsche Bank's 125th anniversary}]_{\text{DOBJ.}}. \ (\text{ETrans\_SHARE\_009}) \]

• a higher frequency of verbs taking certain types of prepositional object in German than in English
Crossing Lines between words and grammatical functions

- using of different tenses:
  - main verbs in German => auxiliar verbs in English

(a) And what has [happened] \(_{PRED}\) before a few years have passed? (EO_FICTION_006)
(b) Und was [geschieht] \(_{FIN}\), ehe noch ein paar Jahre vergangen sind? (GTrans_FICTION_006)
Crossing Lines between words and grammatical functions

• main verbs in German => negation using auxiliar verbs in English
  (main verbs as non-finite verbs)

\[ a. \quad \text{Aber Sie [wissen]}_{\text{FIN}} \text{ nichts. (GO_FICTION_007)} \]
\[ b. \quad \text{But you don't [know]}_{\text{PRED}} \text{ anything. (ETrans_FICTION_007)} \]
## Crossings

<table>
<thead>
<tr>
<th>FICTION</th>
<th>SHARE</th>
<th>SPEECH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E2G</strong></td>
<td><strong>G2E</strong></td>
<td><strong>E2G</strong></td>
</tr>
<tr>
<td>dobj → subj</td>
<td>probj → dobj</td>
<td>compl → probj</td>
</tr>
<tr>
<td>compl → dobj</td>
<td>dobj → subj</td>
<td>dobj → subj</td>
</tr>
<tr>
<td>subj → dobj</td>
<td>fin → pred</td>
<td>dobj → probj</td>
</tr>
<tr>
<td>dobj → fin</td>
<td>compl → subj</td>
<td>compl → dobj</td>
</tr>
<tr>
<td>dobj → probj</td>
<td>subj → dobj</td>
<td>dobj → compl</td>
</tr>
<tr>
<td>fin → dobj</td>
<td>dobj → compl</td>
<td>compl → subj</td>
</tr>
<tr>
<td>adv_mod → dobj</td>
<td>fin → compl</td>
<td>probj → dobj</td>
</tr>
<tr>
<td>pred → fin</td>
<td>pred → fin</td>
<td>subj → dobj</td>
</tr>
<tr>
<td>compl → subj</td>
<td>fin → subj</td>
<td>fin → pred</td>
</tr>
<tr>
<td>adv_cause → dobj</td>
<td>fin → dobj</td>
<td>pred → fin</td>
</tr>
</tbody>
</table>

- most frequent crossing lines
Preliminary study

- random 300 sentence pairs
- both translation directions
- register FICTION, SHARE, SPEECH (50 sentences each)
  - E2G_FICTION
  - E2G_SHARE
  - E2G_SPEECH
  - G2E_FICTION
  - G2E_SHARE
  - G2E_SPEECH
Results

<table>
<thead>
<tr>
<th></th>
<th>OK+KP+FVG +PH</th>
<th>OK-PW</th>
<th>Anteil OK in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2G_FICTION</td>
<td>24</td>
<td>16</td>
<td>66,67</td>
</tr>
<tr>
<td>E2G_SHARE</td>
<td>31</td>
<td>13</td>
<td>41,94</td>
</tr>
<tr>
<td>E2G_SPEECH</td>
<td>28</td>
<td>21</td>
<td>55,26</td>
</tr>
<tr>
<td>G2E_FICTION</td>
<td>30</td>
<td>20</td>
<td>66,67</td>
</tr>
<tr>
<td>G2E_SHARE</td>
<td>32</td>
<td>15</td>
<td>46,88</td>
</tr>
<tr>
<td>G2E_SPEECH</td>
<td>35</td>
<td>19</td>
<td>54,29</td>
</tr>
</tbody>
</table>

• verb-to-verb parallelism without change of perspective (semantic differences)
Feasibility study

• RBMT System – CAT2 (Haller 1993; Sharp 1994)
  • free available
  • Windows operating system
  • grammar revised with an text editor
  • developed as a side line of EUROTRA

• analysis:
  • morphologic layer
  • syntactic layer
  • interface structure => transfer
Setting rules

• ‘Unser Umsatz beträgt eine Milliarde.’

atom =
{lex=be,frame={arg1:
arg1={semf=abs_economy_unit},
arg2={role=attr}}}}.[]
<=>
{lex='beträgen',frame={arg1:
arg1={semf=abs_economy_unit}}}].[].
atom =
{lex=be,frame={arg1:
arg1={semf~=abs_economy_unit}}}].[]
<=>
{lex='sein',frame={arg1:
arg1={semf~=abs_economy_unit}}}].[].
Setting rules

• ‘Mich beunruhigt die vom Himmel fallende Asche.’

\[
\text{worry}_\text{beunruhigen} = \\
\{\text{role}=\text{proposition}\}. \{\text{cat}=v, \text{lex}=\text{worry}\}, \\
\text{np1}:\{\text{focus}=1, \text{case}=\text{nom}\}, \text{np2}:\{\text{focus}~\approx 1\} \\
\leq \\
\{\text{role}=\text{proposition}\}. \{\text{cat}=v, \text{lex}=\text{'beunruhigen'}\}, \\
\text{np2, np1}.
\]

• NP1 and NP2 switching positions (die vom Himmel fallende Asche / the ash falling from the sky)
Setting rules

• ‘Mich beunruhigt die vom Himmel fallende Asche.’

German: NP, PP stands on the left side of the participle (die vom Himmel fallende)

English: NP, PP stands on the right side of the participle (falling from the sky)
Setting rules

- ‘Mich beunruhtigt die vom Himmel fallende Asche.’

- transfer to the interface structure

**German**

```
partp =
{cat=partp}.
[ nppp:{cat=(np;pp)},
  part:{cat=(part1;part2)}]
<=>
{role=mod,cat=(part1;part2)}. [ part:{cat=(part1;part2)},
  nppp:{cat=n} ].
```

**English**

```
partp =
{cat=partp}.
[ part:{cat=(part1;part2)},
  nppp:{cat=(np;pp)}]
<=>
{role=mod}. [ part:{cat=(part1;part2)},
  nppp:{cat=n} ].
```
Setting rules

- ‘Mich beunruhigt die vom Himmel fallende Asche.’

- position of a participial phrase

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ np = {cat=np}].[^{cat=det},^{cat=(partp;ap)},{cat=n}].</td>
<td>[ np = {cat=np}.[^{cat=det},^{cat=ap},{cat=n},^{cat=partp}]].</td>
</tr>
</tbody>
</table>

[^ - optional element
* - 0-n any elements of a kind]
Setting rules

• ‘Mich beunruhigt die vom Himmel fallende Asche.’

• participal phrase before and after adjective

**German**

```
np =
    {cat=np}. [ 
        ^a:{cat=ap},
        ^part:{cat=partp} ,
        n:{cat=n} ]
<=>
    {d=no}. [ 
        n:{cat=n},
        ^part:{cat=(part1;part2)},
        ^a:{cat=a} ].
```

**English**

```
npnodefsing =
    {cat=np,type=T,type~==rel}. [ 
        *a:{cat=ap},
        n:{cat=n,type~==pron},
        ^part:{cat=partp} ]
<=>
    {d=indef,type~==(pron;rel)}. [ 
        n:{cat=n},
        ^part,
        *a:{cat=a} ].
```
Results

• C1 – CAT2 without implemented rule set
• C2 – CAT2 with implemented rule set
• GT – translation with Google Translate

Unser Umsatz beträgt eine Milliarde.
C1: Beträgt our coverage a billion.
C2: Our coverage is a billion.
GT: Our turnover is one billion.

Our coverage is a billion.
C2: Unser Umsatz beträgt eine Milliarde.
GT: Unsere Berichterstattung ist eine Milliarde.
Results

• C1 – CAT2 without implemented rule set
• C2 – CAT2 with implemented rule set
• GT – translation with Google Translate

Mich beunruhigt die vom Himmel fallende Asche.
C1: I worries who sky falling ash.
C2: The ash falling from sky worries me.
GT: I am disturbed by the ash falling from the sky.
References

Thank you for your attention!