METAL

Mechanical Translation and Analysis of Languages
• Machine translation research at the Linguistics Research Center, Texas began in 1970 with Rolf A. Stachowitz as principal researcher. Research funding came from the USAF Rome Air Development Center and other U.S. government agencies for the Metal system.

• Siemens AG began funding system in 1979.

• For market testing, first operational version of the system transmitted in 1985.
• fully-automatic high quality rule based machine translation system

• attempted Interlingua experiments, but it essentially adopted a transfer approach.

• besides the German to English system which has been operational for quite some time, German to Spanish, Dutch to French, English to German have been released
System structure

• The linguistic component of Metal is developed in Lisp and the text processing component are written in C.

• The hardware package of system consist of several translator workstations and a dedicated Lisp machine running as a server in the background. This server linked via Ethernet to a multi-user translator workstation. From these terminals, translation jobs are started and all the tasks are handled.

• The Metal system was built in a highly modular way. There is a language independent linguistic processor that includes language-specific modules for analysis, transfer and synthesis.
Translation Process

**SINIX System**
- Text Acquisition

**Deformatting Programs**

**AI System**
- Pre-analysis Programs

**Intercoder**
- Translation Programs
  - Analysis
  - Transfer
  - Synthesis

**Reformatting Programs**
- Postediting

**Text Output**

- Text transfer via data link or input facilities (floppy disk, magnetic tape, page reader)
- Separation of language and format data
- Processing of special formats (diagrams, tables)
- Generation of word lists for lexicon coding
- Interactive expert system for lexicon update
- Translation
- Merging of language and format data
- Revision of translation
- Word processing system
  - Printer output
  - Typesetting
Lexical Analysis

• Function words (prepositions, conjunctions, pronouns, determiners etc.), general words, technical words, prefixes, infixes, inflectional endings, punctuation, symbols constitutes the lexical database for German and English entries.

• The Metal System contains lexicons which are either monolingual or bilingual. In analysis or generation, system used monolingual lexical entries that includes the canonical form.

• The bilingual lexicons reveal correspondence between source and target language.
German monolingual lexicons

- For the English monolingual lexicons, the principals and particular lexical entries are the identical with the German.
• (VST): the dictionary form *gehen* is a verb stem
• (ALOs): stem forms or allomorphs of *gehen*; *gang*, *ging* and *geh*
• (PLC (WI)): *ging* and *geh* which ALOs of *gehen* must be preceded by a blank. (PLC (NF NI)): non-blank characters, the past participle marker and an ending, must precede and follow the ALO *gang*.
• (TAG (ALL)): *gehen* is a general term
• (SNS): a unique system-assigned number for each entry
• (PX (NIL)): *gehen* takes no prefix as canonical form
• (ARGS): *gehen* appears either with no non-subject or with a single prepositional phrase argument (PP)
• (TT (I)): transitivity type feature showed that *gehen* is intransitive
• (AX (sein)): auxiliary of *gehen* is *sein*
• (CL): inflectional classes defines various allomorphs, such as PP-GEEN show that *gang* as allomorphs of *gehen* takes a *ge*–participle and an –*en* ending.
Bilingual lexicons

• There is an example for the German verb *gehen* with transfer entry;

\[
\begin{align*}
& (\text{go} \quad (\text{gehen}) \ \text{VST} \ (\text{CAT VST}) \ (\text{PX NIL}) \ (\text{PF FIN INF PAPL})) \\
& (\text{outgo} \quad (\text{gehen}) \ \text{VST} \ (\text{CAT VST}) \ (\text{PX NIL}) \ (\text{PF PRPL}))
\end{align*}
\]

• According to first entry, *gehen* is equivalent to the English *go* word under some certain conditions such as the German stem doesn’t have a prefix (PX), also its predicate form (PF) is infinitival (INF), finite (FIN) or past participial (PAPL). The second entry shows that gehen is outgo while it is a present participle (PRPL) and has no prefix.
Grammar Rule composition

- Words, phrases, clauses are built by grammar rules which consist of a number of parts.

- VB -> VST V-FLEX rule:

  First line that defines which the rule applies, a verb (VB) may include the verb stem (VST) and the verb ending (V-FLEX).

  Third line defines the limit of this sample rule, so the verb stem must be preceded by a blank (REQ WI). On the contrary, the verb ending mustn’t be preceded by a blank (NRQ WI). TEST part shows the validity and constraints of the grammar rule.

```
VB   VST   V-FLEX
0    1     2
--   (REQ WI)  (NRQ WI)
TEST (INT 1 CL 2 CL)
   (OR (NOT (INT 2 PF PAPL))
      (INT 1 CL PP-T PP-ET PP-EN PP-N))
   (OR (INT 2 PF PAPL INF)
      (RET 2 WF))
CONSTR (CPX 1 ALO CL)
     (CPY 2 PS NU TN MD PF WF)
     (ADD WI)
     (AND (INT 2 PS 3)
        (INT 2 NU SG)
        (PRF 2))
     (*TAG ALL)
TRANSF (XFM VB)
```
• The appropriate syntactic tree is constituted using right hand elements as the children of left hand elements by the CONSTR part of the rule, after tests are succeeded.

• CONSTR builds the syntactic tree for our sample rule, with copying ((CPX), (CPY)) and adding ((ADD), (PRF)) crucial information to the new node. (TRANSF) Transfer applies a transformation named (VB).

```
   VB
   |
----------------------
|                       |
VST                  V-FLEX
```
Grammar Rule operation: Analysis

(gehen  CAT (VST)
 ALO  (geh)
 PLC  (NF NI)
 TAG  (ALL)
 SNS  (3)
 PX  (NIL)
 ARGS  (#PP #NIL)
 TT  (I)
 AX  (sein)
 CL  (IMP-1 INF-EN PRI-1 PRS-1)
)

(tl  CAT (V-FLEX)
 ALO  (t)
 PLC  (NI)
 SNS  (?)
 CL  (PRI-1 PRI-2 PRI-4 PRI-5 PRI-6 PRI-16)
 NU  (PL)
 PS  (2)
 MD  (IND)
 PF  (FIN)
 TN  (PR)
)

(tl  CAT (V-FLEX)
 ALO  (T)
 PLC  (NI)
 SNS  (8)
 CL  (PRI-1 PRI-2 PRI-4 PRI-5 PRI-11 PRI-13 PRI-14)
 NU  (SG)
 PS  (3)
 MD  (IND)
 PF  (FIN)
 TN  (PR)
)
Rule operation: Transfer

$nach \text{ 3 Stunden geht die Ausgabe auf Magnetband}$

$the \ output \ goes \ to \ magnetic \ type \ after \ 3 \ hours.$
Figure 15.3 InterCoder screen with pop-up window showing possible inflection paradigms
Exhibit A:
an original German text

CSE Sprachgabe-Geräte
Einführungsschrift

1 Einleitung

2 Technik der Spracherkennung
2.1 Allgemeiner Überblick
Practical Consideration

- According to five years experimenting, correctness varied from 45% to 85% for the Metal system.
- This system indicated an average performance about 2+ seconds per unit word with Symbolics 3600 Lisp Machine, 512K and 36-bit words of physical memory.
- Improvements after particular objections;
  - coverage of the grammar was extended
  - the de-formatting and reformatting programs were redesigned
  - errors in the general system lexicon were corrected.