

Corpora in a systemic functional approach to translation teaching

Robert Spence
Saarbrücken

1. Description
2. Theoretical background
3. Aims
4. Error classification
5. Higher-level errors

1. Description

1.1. When, where, what, how many

- 1988–1992 (KM)U Leipzig
- undergraduate translation students
- weekly translation exercises German–English
- student profiles:
 - L1 German
 - L2 English
 - L3 Russian, French, Spanish or Arabic
- 49 source texts:
 - popular science
 - economics
 - politics
 - tourism
 - literature
- 1000 target texts
- 250,000 words
- texts to be discussed:
 - WELTBEVO(elkerung) (100)
 - SITZUNGS(saal) (37)

1.2. Alignment by ‘chunking’: labelling

for display purposes:

vertical alignment of sentence-sized chunks

[45.01.02.000] Nachdem im Juli 1987 ...

[45.01.02.001] After the world population ...

[45.01.02.003] After in July 1987 ...

[45.01.02.004] After another baby ...

Version No.

Sentence No.

Paragraph No.

Text No.

1.3. ‘Chunking’: does it work?

[45.01.02.000]

Nachdem in Juli 1987 —
statistisch durch die Geburt eines Kindes in Jugoslawien —
die Weltbevölkerung eine Zahl von 5 Milliarden Menschen erreicht hatte,
prognostizieren die Experten des Bevölkerungsfonds der UNO
eine weitere Zunahme.

[Literal translation]

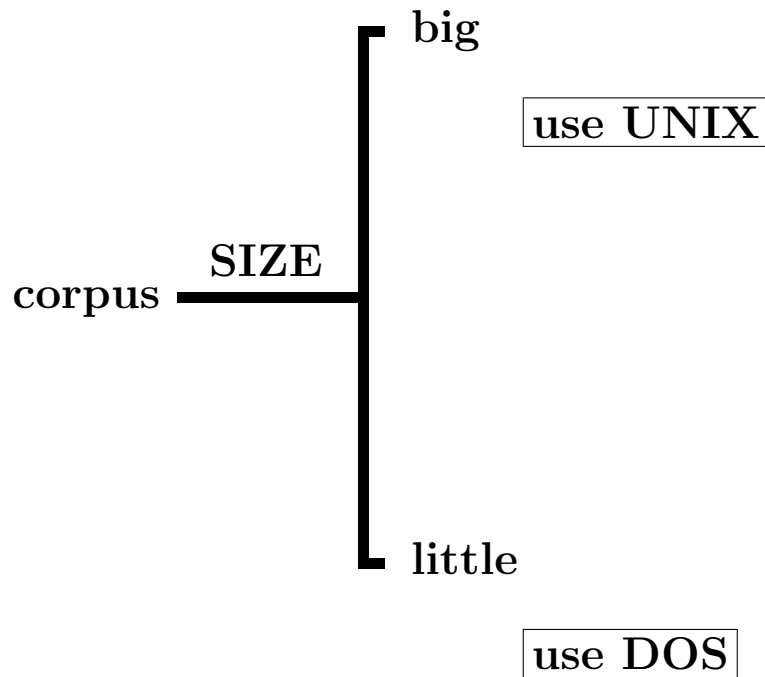
After in July 1987 —
statistically by the birth of a child in Yugoslavia —
world population (had) reached a figure of 5 billion people,
the experts of the United Nations Fund for Population Activities predict a further increase.

[45.01.02.005]

In July 1987 the population of the world reached 5 billion people.
Formally this event was marked by the birth of a child in Yugoslavia.
Now experts from the United Nations Demographic Fund forecast a further increase.

1.4. Storage

Which OS?



Which WP? (Which version?)

- Word,
- AmiPro,
- WordPerfect,
- LocoScript,
- ...

1.5. Word processing

L^AT_EX

Advantages:

- standard ASCII characters
- easy conversion to SGML / HTML
- separation / optimization of
 - display format
 - storage format

2. Theoretical background

2.1. Characterization

SFL_x

- Language (and linguistics) as
 - comprehensive
 - extravagant
 - indeterminate *
 - non-autonomous
 - variable
 - * but formalizable (at a price...)

- Language as ‘meaning potential’
(formalized as paradigmatic SYSTEM)

- Text as ‘instantiation of meaning potential’
(formalized as syntagmatic STRUCTURE)

- Sociological perspective:
Text(= sociosemantic behaviour)-in-situation

2.2. Instantiation and realization

I N S T A N T I A T I O N

CULTURE

SITUATION

R
E
A
L
I
Z
A
T
I
O
N

LANGUAGE

TEXT

3. Aims

3.1 Investigation of “Learner English”

- L1 interference
- L3 interference?

3.2 Didactic applications

3.3 Computational applications

3.1 “Learner English”

Basic notion:

— Interference as a ‘mapping’ of L1 (L3?) onto L2

Ideology:

— a non-cognitivist approach

Method:

— statistical studies of behaviour of populations

Advantage:

— allows sidestepping of dichotomy:

‘errors’ (competence)	vs.	‘mistakes’ (performance)
occur repeatedly in corpus		occur sporadically corpus

3.2 Didactic applications

a.) Problem: incomprehensibility

MODE = written to be read silently,
but nevertheless read aloud

b.) Bad solution: “sea of paper” syndrome

MODE = written
(Ss scanning while listening)
TENOR = teacher-centred

c.) Better solution: less paper, more thinking

- (i) reduce amount of data
- (ii) allow self-correction

d.) Optimal solution: more technology, less control

MODE = on-line
TENOR = student-centred (“hands-on”)

3.3 Computational applications

(even more technology, even less control)

Goals:

- Integration with text generation system (PENMAN)
- Development of interactive learning systems

Desiderata:

- sentence generators (NIGEL + KOMET): optimal form of representation of system networks (hypertext-based? spreadsheet?)
- L1 as metalanguage
- “choosers” or “choice experts” (PENMAN)
- realization statements: immediate operation
- parser for back-checking

4. Error classification

4.1. A cosy clutter of categories ...

addit	ambig	art	
awkward	clumsy		coll
constr	context	def art	
expr	gr	GR	
grrr	GRRR!	idiom	
indef art	logic	logical sequence	
meaning	modal	om	
orth	overtr	prep	
punct	ref	shift of emphasis	
	sstr	SSTR	
stress	style	superfl op	
T	TA	T seq	
text	tr	undertr	
Vb	VP	voice	
word	w.o.		

... and of combinations thereof

expr/tr		expr/style/logic	
	gr/tr	tr/ambig	tr/stress
w/context		w/tr	

4.2. What's wrong with the cosy clutter

No theory of LANGUAGE

- no mutual 'definition' of categories

No theory of TEXT

- two occurrences of 'same' error?
or
one occurrence each of two 'separate' errors?
- no basis for objective appraisal of 'seriousness' of errors
(= impact on usability of translation)

4.3. A system(at)ic classification

based on a theory of language

based on a theory of text

Classification dimensions:

a.) by origin in system

stratum

rank

metafunction

b.) by degree of seriousness

4.4. Classification by STRATUM

CONTEXT

(interface to extralinguistic situation)

SEMANTICS

(context-sensitive meaning)

LEXICOGRAMMAR

(context-neutral meaning)

GRAPHOLOGY

(= emic)

GRAPHETICS

(= etic; concrete substance)

4.5. Classification by RANK

e.g. at the lexicogrammatical stratum:

grammar:

clause complex

clause

group/phrase complex

group/phrase

word complex

word

morpheme complex

morpheme

lexis:

collocation

lexical item

4.6. Classification by METAFUNCTION

e.g. at clause or clause complex rank in lexicogrammar:

ideational

 logical

 experiential

interpersonal

textual

4.7. Classification by SERIOUSNESS

3 seriously wrong

2 wrong

1 just on the wrong side of the
borderline of acceptability...

0 just on the right side of the
borderline of acceptability...

4.8. Examples (1)

[45.02.07.000]

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen; inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen und hat sich damit in nur 25 Jahren mehr als verdoppelt.

[45.02.07.073]

In 1960, for example, 276 million people lived on the African continent.

Till 1985 the population grew to 553 million.

This way it was more than doubling within 25 years only.

Ser: Str: Rnk: Mfn: Problem:

3 gr cl exper circ: durative temporal,
with

non-durative process

Produced:

Till 1985 the population grew to 553 million

ShouldBe:

By 1985 the population had grown to 553 million /

The population grew to 553 million by 1985

4.9. Examples (2)

Ser: Str: Rnk: Mfn: Problem:
 2 gr cl textu conj Adj: This way
 for
 Thus
 from
 daher

Produced:

This way it was more than doubling

ShouldBe:

, and thus more than doubled /
 , thus more than doubling

Ser: Str: Rnk: Mfn: Problem:
 3 gr cl exper tense: [a 0] [b -]
 for
 [a -]
 from
 [Imperfekt]

Produced:

(it) was more than doubling within 25 years only

ShouldBe:

(it) more than doubled in only 25 years /
 more than doubling in only 25 years

5. Higher-level errors

5.1. Stratifying ‘context’

IDEOLOGY

GENRE

REGISTER

(Field, Tenor, Mode)

SEMANTICS

(context-sensitive meaning)

LEXICOGRAMMAR

(context-neutral meaning)

GRAPHOLOGY

(= emic)

GRAPHETICS

(= etic; concrete substance)

5.2. Feature selection expressions

(output of systemic parser)

Tabulate these ‘distinctive features’ by clause:

01 past material: receptive temporal-circ
(.. was destroyed .. in 1941)

02 past material: receptive temporal-circ
(.. was opened .. in 1950)

...

...

21 present relational
(.. is .. a replica of)

22 present relational
(.. is .. an enlargement of)

...

...

5.3. Generic structure (1)

‘microregisterial variation’ within a text:

narrative and descriptive genres in a ‘single’ text

01 tense: past circumstantiation: temporal

02 tense: past circumstantiation: temporal

03 tense: past circumstantiation: spatial

04 tense: past circumstantiation: temporal

..

..

21 tense: present circumstantiation: spatial

22 tense: present circumstantiation: spatial

23 tense: past circumstantiation: spatial

24 tense: present circumstantiation: spatial

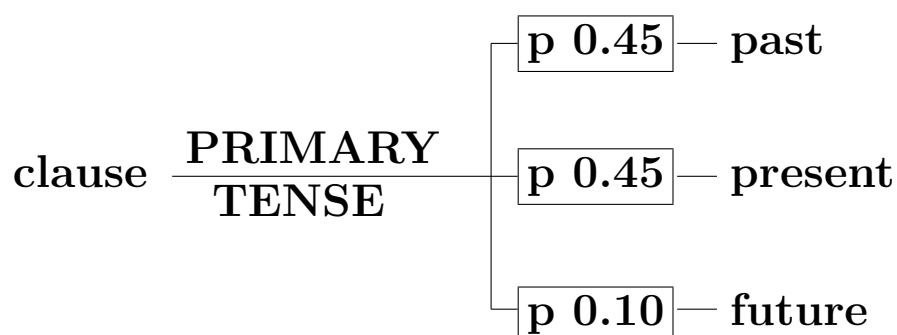
..

..

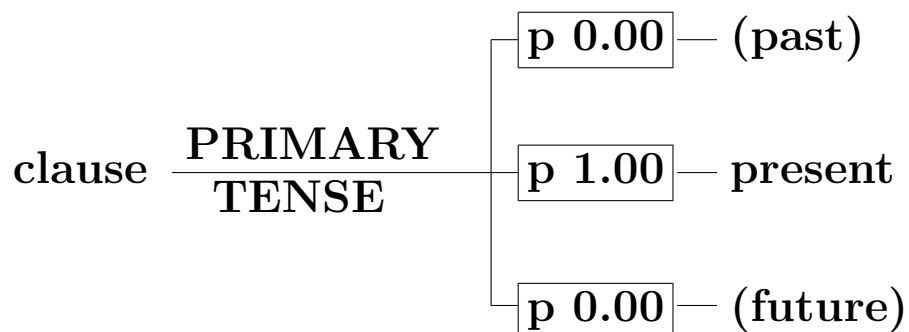
5.4. Register

realization of register (text type) features via re-weighting of inherent (context-neutral) probabilities:

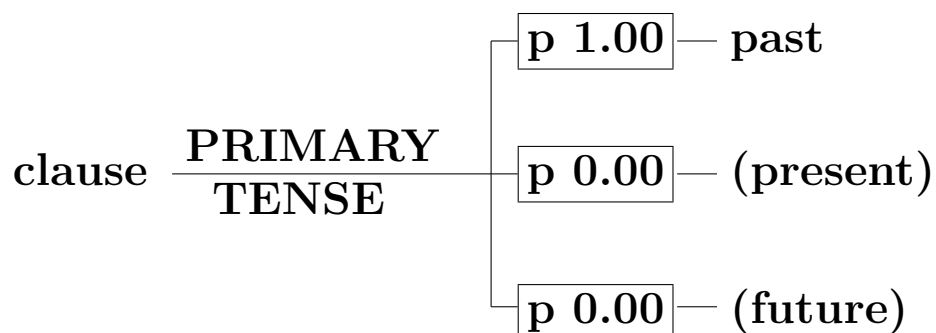
RESTING STATE:



REGISTER A:



REGISTER B:



5.5. Generic structure (2)

nature and sequence of 'parts' of texts

Genre: history

past development

(what was it like originally? and what happened?)

+

present description

(and what is it like now?)

Genre: tourist guide

present description

(where am I? where/what/how is everything?..)

+

past development

(.. and why is it like that?

= how did it come to be that way?)

5.6. Genre constraining Register

GENRE:

There is no such thing as a university lecture on bicycle maintenance.

REGISTER:

<u>FIELD</u>	[...
		bicycle maintenance
]	...
<u>TENOR</u>	[...
		professor-to-students
]	...
<u>MODE</u>	[...
		lecture
]	...

5.7. Ideology constraining Genre

Little bureaucrat reporting:

Mit der vollen Wirkung der Wirtschafts- und Währungsunion in Deutschland mußte auch unser Unternehmen neue Marktstrategien entwickeln, vor allem durch die Einbrüche im traditionellen osteuropäischen Markt.

The full effect of the Economic and Monetary Union in Germany has forced us to develop new market strategies to compensate for the sharp tumble in sales to our traditional Eastern European customers.

Little battler boasting:

German Economic and Monetary Union hit many enterprises hard. Ours was no exception. Reacting quickly to the sharp tumble in sales to Eastern Europe, we successfully developed new market strategies to compensate for the loss of traditional customers.