A Corpus of Student L1–L2 Translations

Robert Spence
Universität des Saarlandes

1. Objections

2. Issues

3. Description of corpus

4. Theoretical framework

5. Applications

1. Objections

Why?!?

TS skepticism:

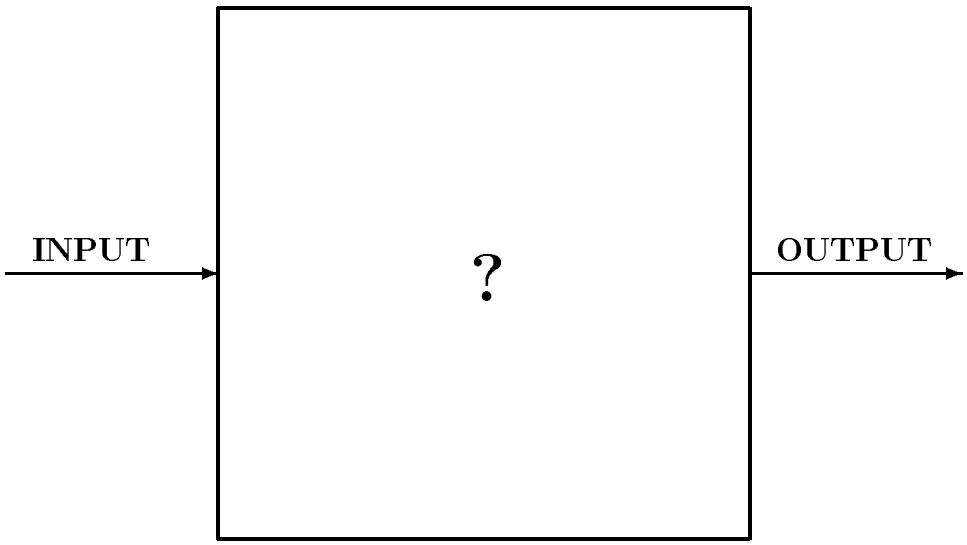
- professional L2–L1 translations = 'good'
  (& therefore 'useful', e.g. translation memory systems)
- student L1–L2 translations = 'bad'
  ("junk in, junk out"?)

FLT skepticism:

- instrumentalization of disciplines?
  - translation as a component of FLT?
  - FLT as a part of translator training?
- degree of constraint on learner productions
- nature and sources of aberrancies

2. Issues

- variation and constraint
  - interference/interlanguage (IL) hypothesis



  - multilingual text generation parallels
  - machine translation (MT) parallels

- grammatical functionalism
  - non-exaggeration of non-standardness
  - "remnants of L1 functionality"?
  - interorganism perspective

2. Issues (contd.)

- **instantiality**
  - errors ("non-standard wiring"; "IL")
  - mistakes ("random misfirings")
  - instantial interference?

- **L1-independence?**

- **"per genus et differentiam"**
  - L2 English?
  - English L2?

  (what about L3??)

- **a tangled braid ...**
  - non-nativeness [TL = L2, L3, ...]
  - un-Englishness [(T)L $\neq$ English]
  - translationese [SL $\longrightarrow$ TL]

3. Description of corpus

- 1988–1992
- University of Leipzig
- undergraduate students with
  - major in Translation
  - minor in Translation
- translations done as
  - weekly exercises
  - final examinations
- SL German $\longrightarrow$ TL English
- student profiles:
  - L1 German
  - L2 English
  - L3 Russian, French, Spanish or Arabic
  - L1 English for comparison
- 49 source texts:
  - popular science
  - economics
  - politics
  - tourism
  - literature
- 1232 target texts
- ca. 280,000 words

No.	Filename	Translations			No.	Filename	Translations		
		S	N	T			S	N	T
48	WELTBEVO	105	2	107	40	STEINZEI	12	1	13
34	SCHAERFE	99	1	100	42	THUERING	11	1	12
04	BARTHELS	90	2	92	15	FORCLAZ	9	2	11
36	SITZUNGS	74	2	76	23	KINDKREB	10	1	11
46	VORWORT	60	3	63	37	SKEPSIS	10	1	11
35	SCHUHE	43	2	45	39	SPRACHEN	10	1	11
27	MASCHE	43	1	44	30	OZRAKETE	9	1	10
18	GRUENA	35	4	39	49	WESTLOHN	9	1	10
08	BRUECKEN	36	2	38	19	GRUSSWOR	7	2	9
10	COMPKRIM	37	1	38	50	WIEDERBE	8	1	9
12	DUFT	37	1	38	31	PAPIER	7	1	8
24	LAERM	37	1	38	38	SPORTMED	7	1	8
16	GEBAEREN	33	2	35	47	WAEHRUNG	6	1	7
20	HAELFTE	34	1	35	22	HOEHEPUN	5	1	6
13	EUREKA	29	1	30	11	COSPUDEN	4	1	5
28	MEERESOR	29	1	30	26	LEISTUNG	4	1	5
17	GOLDRAUS	25	2	27	44	VERTRAUE	4	1	5
51	WUENSCHE	25	1	26	09	BYZANTIN	3	1	4
14	FAHRPLAN	24	1	25	33	REINLUFT	3	1	4
03	ANGEBOT	22	1	23	06	BEITRAGS	2	1	3
25	LEINE	21	1	22	07	BERLINER	2	1	3
05	BASIC	19	2	21	45	VIELFALT	2	1	3
21	HERZCHIR	17	1	18	02	ALLES	1	1	2
01	AIDS	15	1	16	29	OSSIJOBS	1	1	2
43	TWAIN	14	2	16	32	RAUMKATZ	1	1	2
41	TEMP	13	1	14	52	ZUFALL	1	1	2
TOTAL NO. OF TRANSLATIONS:							1164	68	1232

S = Students, N = Native speakers, T = Total

Alignment & labelling; text "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 ...

"Chunks" and "translation units"?

[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.

Error classification

[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 (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

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: [α -] [β 0] for [α -] 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

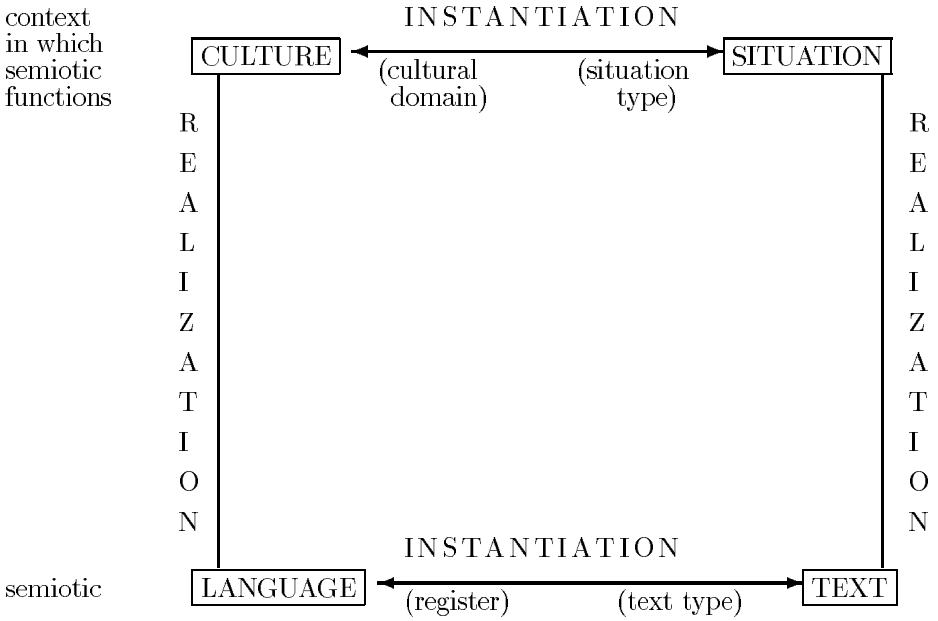
4. Theoretical framework

Systemic Functional Linguistics (SFL)

- Disadvantages
  - cultural foreignness
  - terminology
  - high and special demands
  - orientation to low-morphology languages

- Advantages
  - applicational versatility
  - rich conceptual apparatus
  - orientation to language use
  - variation-theory-friendliness

- **Language as 'meaning potential'**
  (formalized as paradigmatic **SYSTEM**)
- **Text as 'instantiation of meaning potential'**
  (formalized as syntagmatic **STRUCTURE**)
- **Cultural relativist perspective:**
  Language as non-autonomous w.r.t. culture
- **Sociological/ethnographic perspective:**
  Text(= sociosemantic behaviour)-in-situation



- **Hjemslevian (macrosystemic)**
- **Firthian (microsystemic)**
- **Whorfian (Culture | Language)**
- **Malinowskian (Situation | Text)**

5. Applications

Short term

- **course design**
  - some texts more 'registerially typical' than others?
  - some registers more 'culturally central' than others?
- **error prediction**
- **direct use in class**
  - constraints
    - didactic-theoretical
      - mode (written vs. spoken)
      - tenor (teacher-centredness)
    - logistic

Medium term

- **translation quality assessment (TQA)**
  - error classification
  - error weighting ('error' $\longrightarrow$ 'phenomenon'?)
- **Robustness testing of MT systems**

Longer term

- **computer-aided translation (&/or FL) teaching**