

Einführung Computerlinguistik

Allgemeines zur Linguistik

Hinrich Schütze & Robert Zangenfeind

Centrum für Informations- und Sprachverarbeitung, LMU München

2013-10-14

Take-away

Take-away

- Was ist Linguistik?

Take-away

- Was ist Linguistik?
- Linguistische Grundbegriffe wie z.B. Type und Token

Take-away

- Was ist Linguistik?
- Linguistische Grundbegriffe wie z.B. Type und Token
- Das sprachliche Zeichen

Overview

- 1 Linguistik
- 2 Das sprachliche Zeichen

Outline

- 1 Linguistik
- 2 Das sprachliche Zeichen

Linguistik (Sprachwissenschaft)

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.
- nicht: Programmiersprachen, Logiksprachen, Plansprachen (z.B. Esperanto, Volapük)

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.
- nicht: Programmiersprachen, Logiksprachen, Plansprachen (z.B. Esperanto, Volapük)

Was ist Sprache?

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.
- nicht: Programmiersprachen, Logiksprachen, Plansprachen (z.B. Esperanto, Volapük)

Was ist Sprache?

- Man kann sich ein ganzes Studium lang nur mit dieser Frage beschäftigen.

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.
- nicht: Programmiersprachen, Logiksprachen, Plansprachen (z.B. Esperanto, Volapük)

Was ist Sprache?

- Man kann sich ein ganzes Studium lang nur mit dieser Frage beschäftigen.
- Teile dieser Vorlesung: Die Sicht einer Schule, der [Moskauer Semantischen Schule](#)

Linguistik (Sprachwissenschaft)

Theorie der natürlichen Sprachen

- d.h. natürliche Sprache beschreiben, modellieren
- natürliche Sprache: entstanden in einer historischen Entwicklung, z.B. Dt., Engl., Russ.
- nicht: Programmiersprachen, Logiksprachen, Plansprachen (z.B. Esperanto, Volapük)

Was ist Sprache?

- Man kann sich ein ganzes Studium lang nur mit dieser Frage beschäftigen.
- Teile dieser Vorlesung: Die Sicht einer Schule, der [Moskauer Semantischen Schule](#)
- System von Zeichen (Wortschatz) und Regeln (Grammatik) zur Mitteilung von Bedeutungen

Schulen in der Linguistik

Schulen in der Linguistik

- Strukturalismus

Schulen in der Linguistik

- Strukturalismus
- Semiotik (zweiter Teil der Vorlesung)

Schulen in der Linguistik

- Strukturalismus
- Semiotik (zweiter Teil der Vorlesung)
- Wittgenstein

Schulen in der Linguistik

- Strukturalismus
- Semiotik (zweiter Teil der Vorlesung)
- Wittgenstein
- Kognitive Linguistik

Schulen in der Linguistik

- Strukturalismus
- Semiotik (zweiter Teil der Vorlesung)
- Wittgenstein
- Kognitive Linguistik
- Moskauer Semantische Schule (ist Grundlage eines Teils dieser Vorlesung)

Einige linguistische Grundbegriff

- Token vs. Type
- Wortschatz/Vokabular
- Gebrauch vs. Erwähnung
- Meta- vs. Objektsprache
- Distribution
- Paradigmatische vs. syntagmatische Sprachachse
- Wohlgeformtheit und Grammatikalität
- Deskriptivität vs. Präskriptivität

Token vs. Type (1)

Token vs. Type (1)

Token / Zeichenvorkommnis

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

- Klasse von Token, die von ihren Sprechern und Hörern (bzw. ihren Autoren und Lesern) nicht unterschieden werden und daher als gleich, z.B. als Kopien, wahrgenommen werden

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

- Klasse von Token, die von ihren Sprechern und Hörern (bzw. ihren Autoren und Lesern) nicht unterschieden werden und daher als gleich, z.B. als Kopien, wahrgenommen werden
- ursprüngliche Unterscheidung durch C.S. Peirce

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

- Klasse von Token, die von ihren Sprechern und Hörern (bzw. ihren Autoren und Lesern) nicht unterschieden werden und daher als gleich, z.B. als Kopien, wahrgenommen werden
- ursprüngliche Unterscheidung durch C.S. Peirce

“eine Rose ist eine Rose”

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

- Klasse von Token, die von ihren Sprechern und Hörern (bzw. ihren Autoren und Lesern) nicht unterschieden werden und daher als gleich, z.B. als Kopien, wahrgenommen werden
- ursprüngliche Unterscheidung durch C.S. Peirce

“eine Rose ist eine Rose” → 5 Token,

Token vs. Type (1)

Token / Zeichenvorkommnis

- sinnlich wahrnehmbares, mündlich oder schriftlich geäußertes Zeichen
- konkrete Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, . . .) an einem Ort zu einer Zeit

Type / Zeichentyp

- Klasse von Token, die von ihren Sprechern und Hörern (bzw. ihren Autoren und Lesern) nicht unterschieden werden und daher als gleich, z.B. als Kopien, wahrgenommen werden
- ursprüngliche Unterscheidung durch C.S. Peirce

“eine Rose ist eine Rose” → 5 Token, 3 Types

Token/Type: fuer Wörter oder Wortformen

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:
→ 11 Token,

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:
→ 11 Token, 9 Types

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:
→ 11 Token, 9 Types
- Wörter:

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:
→ 11 Token, 9 Types
- Wörter:
→ 11 Token,

Token/Type: fuer Wörter oder Wortformen

- wichtig ist grundlegende Unterscheidung zwischen 'Wort' und 'Wortform'
Beispiel: "eine Rose ist eine Rose und viele Rosen ergeben einen Strauß"
- Wortformen:
→ 11 Token, 9 Types
- Wörter:
→ 11 Token, 7 Types

Type/token ratio: Written language

<http://www.speech-therapy-information-and-resources.com/downloads/type-token-ratio.pdf>

Type/token ratio: Written language

<http://www.speech-therapy-information-and-resources.com/downloads/type-token-ratio.pdf>

TEXT 1: Written Language

Type/token ratio: Written language

<http://www.speech-therapy-information-and-resources.com/downloads/type-token-ratio.pdf>

TEXT 1: Written Language

But what are thoughts? Well, we all have them. They are variously described as ideas, notions, concepts, impressions, perceptions, views, beliefs, opinions, values, and so on. At times they are brief, coming and going in an instant. On other occasions they seem to endure and we can mull them over again and again in the act we call thinking. We can put them aside, fall asleep, and then return to them later. We refer to them as things we can handle. However, this is just a metaphor.

Type/token ratio: Written language

<http://www.speech-therapy-information-and-resources.com/downloads/type-token-ratio.pdf>

TEXT 1: Written Language

But what are thoughts? Well, we all have them. They are variously described as ideas, notions, concepts, impressions, perceptions, views, beliefs, opinions, values, and so on. At times they are brief, coming and going in an instant. On other occasions they seem to endure and we can mull them over again and again in the act we call thinking. We can put them aside, fall asleep, and then return to them later. We refer to them as things we can handle. However, this is just a metaphor.

→ 87 Token (Wortformen)

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

→ 62 Types

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

→ 62 Types

Type-Token-Ratio =

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

→ 62 Types

$$\text{Type-Token-Ratio} = \frac{\text{AnzahlTypes}}{\text{AnzahlToken}} =$$

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

→ 62 Types

$$\text{Type-Token-Ratio} = \frac{\text{AnzahlTypes}}{\text{AnzahlToken}} = \frac{62}{87} =$$

Type/token ratio: Written language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	we	6	17	asleep	1	33	impressions	1	49	seem	1
2	and	5	18	at	1	34	instant	1	50	so	1
3	them	5	19	beliefs	1	35	is	1	51	the	1
4	are	3	20	brief	1	36	just	1	52	then	1
5	can	3	21	but	1	37	later	1	53	things	1
6	they	3	22	call	1	38	metaphor	1	54	thinking	1
7	to	3	23	coming	1	39	mull	1	55	this	1
8	again	2	24	concepts	1	40	notions	1	56	thoughts	1
9	as	2	25	described	1	41	occasions	1	57	times	1
10	in	2	26	endure	1	42	opinions	1	58	values	1
11	on	2	27	fall	1	43	other	1	59	variously	1
12	a	1	28	going	1	44	over	1	60	views	1
13	act	1	29	handle	1	45	perceptions	1	61	well	1
14	all	1	30	have	1	46	put	1	62	what	1
15	an	1	31	however	1	47	refer	1	TOTAL		87
16	aside	1	32	ideas	1	48	return	1			

→ 62 Types

$$\text{Type-Token-Ratio} = \frac{\text{AnzahlTypes}}{\text{AnzahlToken}} = \frac{62}{87} = 0,713$$

Type/token ratio: Spoken language

TEXT 2: Speech

01 P: so: (.) er: (..) as you were saying about er:: (.)
02 where are you living now Andrew
03 A: Skipton Lodge
04 P: Skipton Lodge?
05 A: mm (...) Skipton Lodge
06 P: yeah (.) do you like it
07 A: yeah I do
08 P: yeah
09 A: I've settled in
10 P: you have (...) good (.) w w what are the things you
11 like about it
12 A: go out in the tow:n
13 P: you go out in the town (...)
14 A: yeah
15 (2.1)
16 with er: Tommy and Martin (.) and er:: (.) Noel
17 P: and?
18 A: NOEL
19 P: oh yes (.) oh he lives there does he?
20 A: yeah he live(s) in the flats
21 P: yeah (.) oh they have flats there do they
22 A: mm
23 (3.3)
24 and er::
25 (2.3)
26 and I went to see (..) (Elaine)

Type/token ratio: Spoken language

TEXT 2: Speech

```
01   P:   so: (.) er: (..) as you were saying about er:: (.)
02       where are you living now Andrew
03   A:   Skipton Lodge
04   P:   Skipton Lodge?
05   A:   mm (...) Skipton Lodge
06   P:   yeah (.) do you like it
07   A:   yeah I do
08   P:   yeah
09   A:   I've settled in
10   P:   you have (...) good (.) w w what are the things you
11       like about it
12   A:   go out in the tow:n
13   P:   you go out in the town (...)
14   A:   yeah
15       (2.1)
16       with er: Tommy and Martin (.) and er:: (.) Noel
17   P:   and?
18   A:   NOEL
19   P:   oh yes (.) oh he lives there does he?
20   A:   yeah he live(s) in the flats
21   P:   yeah (.) oh they have flats there do they
22   A:   mm
23       (3.3)
24       and er::
25       (2.3)
26       and I went to see (..) (Elaine)
```

→ 88 Token

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL	88	
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL	88	
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL		88
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

Type-Token-Ratio =

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL		88
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

$$\text{Type-Token-Ratio} = \frac{45}{88} =$$

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL		88
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

$$\text{Type-Token-Ratio} = \frac{45}{88} = 0,511$$

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL		88
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

$$\text{Type-Token-Ratio} = \frac{45}{88} = 0,511$$

→ etwa gleich viele Token wie in Text 1

Type/token ratio: Spoken language (2)

rank	word	freq	rank	word	freq	rank	word	freq	rank	word	freq
1	yeah	6	13	flats	2	25	as	1	37	to	1
2	you	6	14	go	2	26	does	1	38	Tommy	1
3	and	5	15	have	2	27	Elaine	1	39	've	1
4	in	4	16	it	2	28	good	1	40	went	1
5	the	4	17	like	2	29	living	1	41	were	1
6	do	3	18	lives	2	30	Martin	1	42	what	1
7	he	3	19	Noel	2	31	now	1	43	where	1
8	I	3	20	out	2	32	saying	1	44	with	1
9	Lodge	3	21	there	2	33	see	1	45	yes	1
10	Skipton	3	22	they	2	34	settled	1	TOTAL	88	
11	about	2	23	town	2	35	so	1			
12	are	2	24	Andrew	1	36	things	1			

→ 45 Types

$$\text{Type-Token-Ratio} = \frac{45}{88} = 0,511$$

→ etwa gleich viele Token wie in Text 1

→ reichhaltigeres Vokabular in Text 1

Abdeckung des Wortschatzes durch ein Korpus

Abdeckung des Wortschatzes durch ein Korpus

Type-Token-Ratio:

Abdeckung des Wortschatzes durch ein Korpus

Type-Token-Ratio:

- Verhältnis der Typen zu deren Vorkommen im Text ist Indiz der Abdeckung eines Korpus (in Bezug auf Wörter/Wortformen)

Abdeckung des Wortschatzes durch ein Korpus

Type-Token-Ratio:

- Verhältnis der Typen zu deren Vorkommen im Text ist Indiz der Abdeckung eines Korpus (in Bezug auf Wörter/Wortformen)
- Typischerweise flacht diese Funktion (x-Achse: Token, y-Achse: Types) nach steiler Anfangskurve asymptotisch ab

Abdeckung des Wortschatzes durch ein Korpus

Type-Token-Ratio:

- Verhältnis der Typen zu deren Vorkommen im Text ist Indiz der Abdeckung eines Korpus (in Bezug auf Wörter/Wortformen)
- Typischerweise flacht diese Funktion (x-Achse: Token, y-Achse: Types) nach steiler Anfangskurve asymptotisch ab
- wenn dies eintritt, dann heißt das, dass eine weitere Vergrößerung des Korpus die Abdeckung des Wortschatzes nur geringfügig verbessern wird.

Gebrauch vs. Erwähnung

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

- Ein Wort (oder ein anderes sprachliches Zeichen), mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

- Ein Wort (oder ein anderes sprachliches Zeichen), mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”
- z.B. bzgl. “Hans”:

(3) Es ist fraglich, ob Hans ein guter Vater für das Kind ist.

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

- Ein Wort (oder ein anderes sprachliches Zeichen), mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”
- z.B. bzgl. “Hans”:

(3) Es ist fraglich, ob Hans ein guter Vater für das Kind ist.

Erwähnung (Anführung, *engl.* mention):

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

- Ein Wort (oder ein anderes sprachliches Zeichen), mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”
- z.B. bzgl. “Hans”:

(3) Es ist fraglich, ob Hans ein guter Vater für das Kind ist.

Erwähnung (Anführung, *engl.* mention):

- Ein Wort, über das gesprochen wird, heißt “erwähnt”

Gebrauch vs. Erwähnung

Gebrauch (*engl.* use):

- Ein Wort (oder ein anderes sprachliches Zeichen), mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”
- z.B. bzgl. “Hans”:

(3) Es ist fraglich, ob Hans ein guter Vater für das Kind ist.

Erwähnung (Anführung, *engl.* mention):

- Ein Wort, über das gesprochen wird, heißt “erwähnt”

(4) Es ist fraglich, ob “Hans” ein guter Name für das Kind ist.

Meta- vs. Objektsprache

Meta- vs. Objektsprache

- Metasprache: Sprache, in der man spricht bzw. etwas beschreibt

Meta- vs. Objektsprache

- Metasprache: Sprache, in der man spricht bzw. etwas beschreibt
- Objektsprache: Sprache, über die bzw. über deren Ausdrücke (Zeichen) man spricht

Meta- vs. Objektsprache

- Metasprache: Sprache, in der man spricht bzw. etwas beschreibt
- Objektsprache: Sprache, über die bzw. über deren Ausdrücke (Zeichen) man spricht
- z.B. auf deutsch über die englische Grammatik sprechen

Distribution

Distribution

Distribution eines Zeichens Z:

Distribution

Distribution eines Zeichens Z:

- Verteilung eines Zeichens Z

Distribution

Distribution eines Zeichens Z:

- Verteilung eines Zeichens Z
- Menge der Kontexte, in denen Z vorkommt

Distribution

Distribution eines Zeichens Z:

- Verteilung eines Zeichens Z
- Menge der Kontexte, in denen Z vorkommt
- z.B. Artikel mit Substantiven

Paradigmatische vs. syntagmatische Sprachachse

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

- Beziehung von Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

- Beziehung von Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas
- Ebene der Ersetzung

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

- Beziehung von Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas
- Ebene der Ersetzung

syntagmatische Sprachachse:

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

- Beziehung von Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas
- Ebene der Ersetzung

syntagmatische Sprachachse:

- Beziehung von Zeichen (Wortformen) zu Zeichen in seinem Kontext (in einem konkreten Satz)

Paradigmatische vs. syntagmatische Sprachachse

paradigmatische Sprachachse:

- Beziehung von Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas
- Ebene der Ersetzung

syntagmatische Sprachachse:

- Beziehung von Zeichen (Wortformen) zu Zeichen in seinem Kontext (in einem konkreten Satz)
- Ebene der Kombination

Beispiel zu paradigmatische vs. syntagmatische Sprachachse

(5)

Die Studentin sitzt in der Vorlesung

Ein Student lernt im Seminar

Hans liest im Hörsaal

↑
paradigmatisch
↓

← syntagmatisch →

Beispiel zu paradigmatische vs. syntagmatische Sprachachse

(5)

Die Studentin sitzt in der Vorlesung

Ein Student lernt im Seminar

Hans liest im Hörsaal

↑
paradigmatisch
↓

← syntagmatisch →

Wie könnte man dieses Schema erweitern?

Beispiel zu paradigmatische vs. syntagmatische Sprachachse

(5)

Die Studentin	sitzt	in	der Vorlesung	↑
Ein Student	lernt	im	Seminar	paradigmatisch
Hans	liest	im	Hörsaal	↓

← syntagmatisch →

Wie könnte man dieses Schema erweitern?

- paradigmatisch: *Natascha steht am Fenster*

Beispiel zu paradigmatische vs. syntagmatische Sprachachse

(5)

Die Studentin	sitzt	in	der Vorlesung	↑
Ein Student	lernt	im	Seminar	paradigmatisch
Hans	liest	im	Hörsaal	↓

← syntagmatisch →

Wie könnte man dieses Schema erweitern?

- paradigmatisch: *Natascha steht am Fenster*
- syntagmatisch: *... und denkt über Linguistik nach*

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist
- Noam Chomsky (1957):

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist
- Noam Chomsky (1957):

(6) Colorless green ideas sleep furiously.

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist
- Noam Chomsky (1957):

(6) Colorless green ideas sleep furiously.

vs.

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist
- Noam Chomsky (1957):

(6) Colorless green ideas sleep furiously.

vs.

*Ideas green sleep colorless furiously

Wohlgeformtheit / Grammatikalität

Wohlgeformtheit:

- Ein sprachlicher Ausdruck A aus einer Sprache L heißt wohlgeformt, wenn er (laut Intuition der Sprecher von L) ein gültiger Ausdruck von L ist
- Noam Chomsky (1957):

(6) Colorless green ideas sleep furiously.

vs.

*Ideas green sleep colorless furiously

- nicht wohlgeformte Sätze (Ausdrücke) werden mit Stern gekennzeichnet

Deskriptivität vs. Präskriptivität

Deskriptivität vs. Präskriptivität

deskriptive Theorie:

Deskriptivität vs. Präskriptivität

deskriptive Theorie:

- beschreibt, was der Fall ist

Deskriptivität vs. Präskriptivität

deskriptive Theorie:

- beschreibt, was der Fall ist

präskriptive Theorie:

Deskriptivität vs. Präskriptivität

deskriptive Theorie:

- beschreibt, was der Fall ist

präskriptive Theorie:

- schreibt vor, was der Fall sein soll

Deskriptivität vs. Präskriptivität

deskriptive Theorie:

- beschreibt, was der Fall ist

präskriptive Theorie:

- schreibt vor, was der Fall sein soll

Literatur:

Bußmann, H.: Lexikon der Sprachwissenschaft. Stuttgart 2008

Lewandowski, Th.: Linguistisches Wörterbuch. Heidelberg 1994

Outline

- 1 Linguistik
- 2 Das sprachliche Zeichen

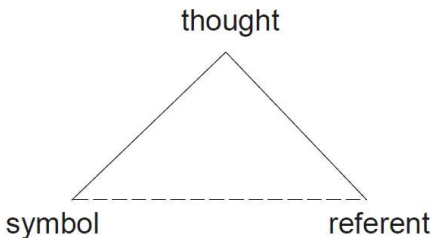
Semiotisches Dreieck

Semiotisches Dreieck

Ogden, Richards: The Meaning of Meaning (1923)

Semiotisches Dreieck

Ogden, Richards: The Meaning of Meaning (1923)



Alternative Bezeichnungen im semiotischen Dreieck

Alternative Bezeichnungen im semiotischen Dreieck

symbol:

Alternative Bezeichnungen im semiotischen Dreieck

symbol:

- signifiant (Saussure)

Alternative Bezeichnungen im semiotischen Dreieck

symbol:

- signifiant (Saussure)
- Signifikant

Alternative Bezeichnungen im semiotischen Dreieck

symbol:

- signifiant (Saussure)
- Signifikant
- Bezeichnendes

Alternative Bezeichnungen im semiotischen Dreieck

symbol:

- signifiant (Saussure)
- Signifikant
- Bezeichnendes

→ Ausdrucksseite des sprachlichen Zeichens

Alternative Bezeichnungen im semiotischen Dreieck

Alternative Bezeichnungen im semiotischen Dreieck

thought:

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)
- Bezeichnetes

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)
- Bezeichnetes
- Sinn (bei Frege)

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)
- Bezeichnetes
- Sinn (bei Frege)
- Bedeutung (außer bei Frege)

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)
- Bezeichnetes
- Sinn (bei Frege)
- Bedeutung (außer bei Frege)
- Intension (Carnap)

Alternative Bezeichnungen im semiotischen Dreieck

thought:

- signifié (Saussure)
- Signifikat (Morris)
- Bezeichnetes
- Sinn (bei Frege)
- Bedeutung (außer bei Frege)
- Intension (Carnap)

→ Inhaltsseite des sprachlichen Zeichens

Alternative Bezeichnungen im semiotischen Dreieck

Alternative Bezeichnungen im semiotischen Dreieck

referent:

Alternative Bezeichnungen im semiotischen Dreieck

referent:

- Denotat (Morris)

Alternative Bezeichnungen im semiotischen Dreieck

referent:

- Denotat (Morris)
- Extension (Carnap)

Alternative Bezeichnungen im semiotischen Dreieck

referent:

- Denotat (Morris)
- Extension (Carnap)
- Bedeutung (bei Frege)

Alternative Bezeichnungen im semiotischen Dreieck

referent:

- Denotat (Morris)
- Extension (Carnap)
- Bedeutung (bei Frege)

→ etwas, das seinen Platz in der außersprachlichen Wirklichkeit hat: Gegenstand oder Ereignis etc., „Ding“

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (1)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (1)

- Bedeutung B eines Zeichens Z (genauer: Ausdrucksseite) ist im Allgemeinen nicht aufgrund von Eigenschaften von Z vorhersagbar (vgl. z.B. *Tisch*)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (1)

- Bedeutung B eines Zeichens Z (genauer: Ausdrucksseite) ist im Allgemeinen nicht aufgrund von Eigenschaften von Z vorhersagbar (vgl. z.B. *Tisch*)
- In der Sprechergruppe hat sich die Konvention (Regel, Übereinkunft) durchgesetzt, Z zu gebrauchen, wenn man B meint (vgl. z.B. Konvention, rechts zu fahren, nicht aber in England)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (1)

- Bedeutung B eines Zeichens Z (genauer: Ausdrucksseite) ist im Allgemeinen nicht aufgrund von Eigenschaften von Z vorhersagbar (vgl. z.B. *Tisch*)
- In der Sprechergruppe hat sich die Konvention (Regel, Übereinkunft) durchgesetzt, Z zu gebrauchen, wenn man B meint (vgl. z.B. Konvention, rechts zu fahren, nicht aber in England)
- Zeichen Z ist (in den meisten Fällen) willkürlich (arbiträr) der Bedeutung B zugeordnet

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*
 - franz. *ouah ouah*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*
 - franz. *ouah ouah*
 - Thai *hoang hoang*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*
 - franz. *ouah ouah*
 - Thai *hoang hoang*
 - japan. *kyankyan*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*
 - franz. *ouah ouah*
 - Thai *hoang hoang*
 - japan. *kyankyan*
 - indones. *gongong*

Arbitrarität und Konventionalität natürlichsprachlicher Zeichen (2)

- Ausnahme von der (völligen) Arbitrarität (aber nicht von der Konventionalität) ist Onomatopöie (Lautmalerei)
 - z.B. Bezeichnung für Gebell von Hunden wird in der Sprache nachgeahmt
 - dt. *wau wau* (Kindersprache auch für Hund)
 - engl. *bow-wow*
 - russ. *gav gav*
 - franz. *ouah ouah*
 - Thai *hoang hoang*
 - japan. *kyankyan*
 - indones. *gongong*
- ist also nicht (bzw. nur sehr wenig) arbiträr, weil am realen Ereignis orientiert (Konvention ist aber dennoch vorhanden)

Take-away

- Was ist Linguistik?
- Linguistische Grundbegriffe wie z.B. Type und Token
- Das sprachliche Zeichen