



UNIVERSITÄT
ZU KÖLN

Computational Linguistics: Big Picture

Einführung in die Informationsverarbeitung

Nils Reiter

October 19, 2023

Brief history of Computational Linguistics I

- ▶ 1950s: DARPA Projects to automatically translate Russian into English
- ▶ 1957/65: Linguistics shifts focus from describing to generating Chomsky (1957, 1965)
- ▶ 1959: Theo Lutz for the first time generates a German poem with a computer Lutz (1959)
- ▶ 1962: Foundation of the “Association for Machine Translation and Computational Linguistics”, 1968 renamed to “Association for Computational Linguistics (ACL)”
- ▶ 1966, ALPAC report: MT more expensive, less accurate and slower than human translation ALPAC (1966)
- ▶ 1968: Foundation of SYSTRAN, first MT company
- ▶ 1975: European commission uses SYSTRAN software (first use of MT on EU level)

Brief history of Computational Linguistics II

- ▶ 1984: First corpus-based commercial MT system Nagao (1984)
- ▶ 1992: Study programs established in Germany (Universities Saarbrücken/Stuttgart)
- ▶ 2011: IBM Watson beats two humans in Jeopardy / Apple Siri launched
https://www.youtube.com/watch?v=WFR310m_xhE
- ▶ 2013: Word embeddings (e.g., word2vec) Mikolov et al. (2013)
- ▶ 2017: Launch of the DeepL Translator
- ▶ 2018: Transformer models: BERT Devlin et al. (2019)
- ▶ 2022: Publicly usable transformer model ChatGPT <https://chat.openai.com>

Embeddings

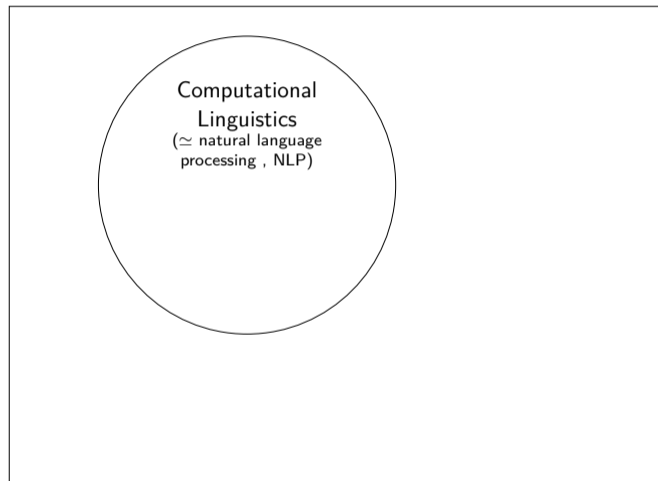
Example (Vector for “köln”)

0.0539 -0.0030 0.0203 -0.1084 -0.0099 0.0705 -0.0546 -0.0433 -0.0096 0.0561 -0.0095 0.0280 0.1726 0.0190 0.0369 0.0217 -0.0002 -0.0309 0.0347 -0.0749
-0.0202 0.0151 -0.0195 0.0001 0.0232 0.0243 -0.0170 -0.0090 -0.0108 -0.0943 0.0376 0.1118 -0.0324 0.0148 -0.0033 0.0537 -0.0681 -0.0733 -0.0201 -0.0329
0.1242 0.0324 -0.0744 -0.0149 -0.0047 -0.0484 -0.0483 0.0481 0.0107 0.0101 -0.0704 0.0500 0.0112 -0.0227 0.0499 -0.0259 -0.0441 0.0712 -0.0157 -0.1271
0.0407 -0.0495 -0.0359 0.0202 0.0024 0.0764 0.0196 0.0267 -0.0117 0.0026 0.0171 -0.0121 -0.1374 -0.0370 0.0247 -0.0113 -0.0094 0.0322 -0.0347 -0.0866 0.0042
-0.0014 0.0067 0.0591 0.0009 0.0085 0.0310 0.0479 -0.0511 0.0198 -0.0886 -0.0274 -0.1364 0.0322 -0.1638 -0.0689 0.0016 -0.1039 0.0059 0.0757 -0.0034 0.1013
-0.0034 -0.0065 -0.0468 0.1577 -0.0065 -0.0478 -0.0004 0.0682 0.0045 -0.0607 -0.0590 0.0343 0.0036 -0.1014 -0.0136 -0.0063 0.0801 0.0360 0.0579 -0.0039
0.0975 0.0500 -0.0558 -0.0095 0.0057 -0.0246 0.1070 -0.0186 0.0669 -0.0781 -0.0569 -0.1286 -0.0834 0.0106 -0.0672 -0.0205 0.0613 0.0290 -0.0545 -0.0481
-0.0882 -0.0489 0.0622 -0.0730 -0.0192 -0.0415 -0.0287 0.0218 -0.0427 -0.0046 0.0255 -0.1164 0.0077 -0.0546 -0.0786 0.0000 -0.0456 0.0943 0.0157 -0.0117
-0.0441 -0.0015 -0.0556 -0.0508 0.0088 0.0418 0.0030 -0.1450 -0.0663 0.0800 0.0172 -0.0289 0.1178 -0.0973 0.0888 0.0637 -0.0295 0.0212 0.0100 -0.0860 0.0035
0.0730 0.0425 -0.0080 0.0885 -0.0166 -0.0765 0.0004 -0.0118 0.0138 -0.0093 -0.0606 -0.0447 -0.0746 0.0131 -0.0447 -0.0763 0.0032 0.1181 0.0542 0.0431
-0.0273 0.0547 0.0135 0.0006 -0.0241 -0.0418 0.0278 -0.0821 -0.0572 -0.0039 0.0214 -0.0196 0.0449 -0.0286 0.0204 0.0681 -0.0901 -0.0266 -0.0287 -0.0874
0.0797 -0.0784 -0.0920 0.0380 0.0411 0.0859 0.0369 0.0595 0.0446 0.0363 -0.0353 -0.0044 -0.0061 0.1134 0.1420 -0.0026 -0.0013 0.0033 0.0508 0.0096 -0.0757
0.0085 -0.0099 -0.0384 0.0218 -0.0259 -0.0112 -0.0212 0.0273 0.0532 -0.0278 -0.0634 0.0317 -0.0022 0.0882 -0.0240 0.0031 -0.0370 0.0747 -0.0097 -0.0315
0.0405 0.0124 -0.1416 -0.0768 0.0363 -0.1248 -0.0134 0.0702 -0.0905 -0.0387 0.0683 -0.0784 0.0886 0.0640 0.0611 -0.0199 -0.0447 -0.1331 -0.1247 0.0540
0.0499 -0.0212 -0.0544 -0.1161 -0.0729 0.0894 0.0532 0.0164 -0.0039 -0.0108 -0.0248 -0.1021 -0.0549 -0.0318 0.0309 -0.0691

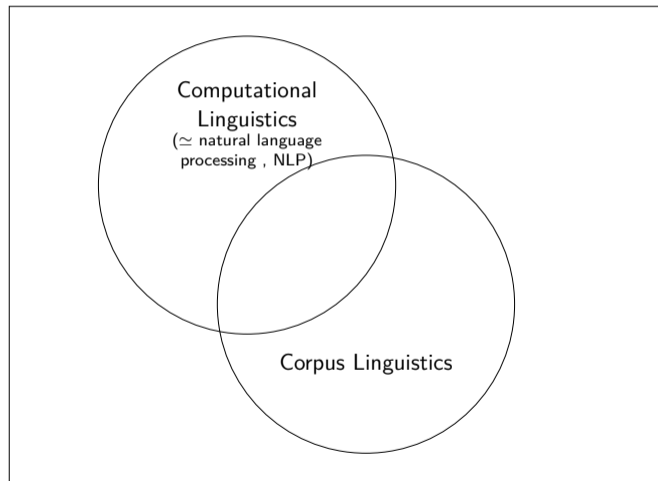
Section 1

Computational Linguistics

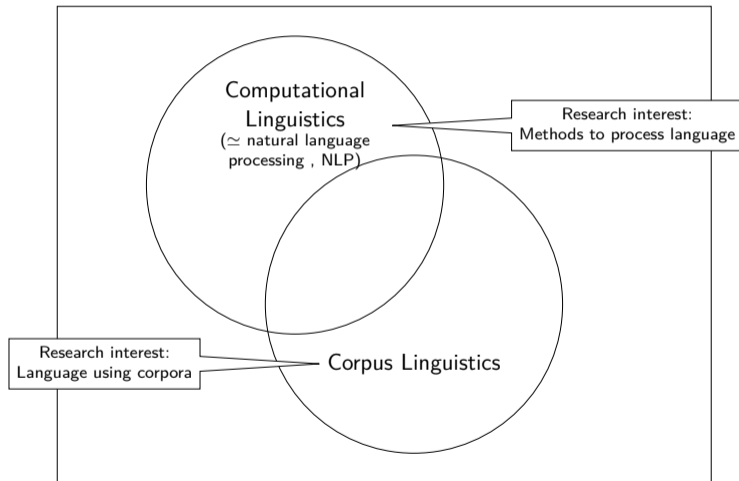
Disciplinary Placement



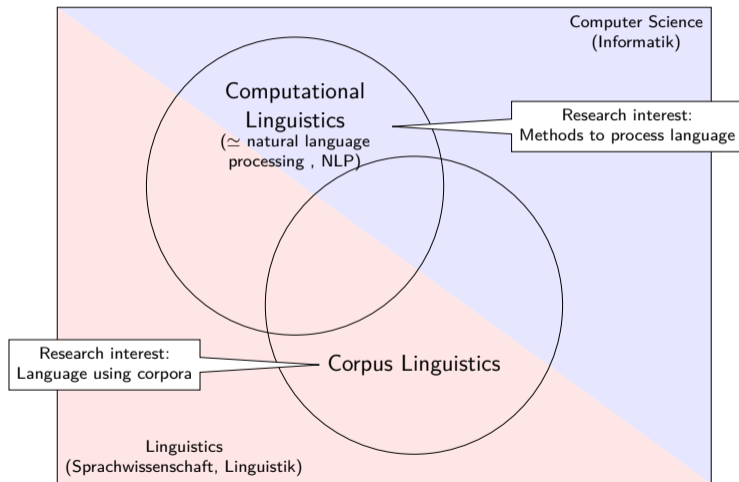
Disciplinary Placement



Disciplinary Placement



Disciplinary Placement



Digital Humanities and Computational Linguistics


Today

- ▶ Digital Humanities, broadly: Working with 'digital methods' on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area
 - ▶ ... but this is a minority position in CL, often also seen as part of AI

Reiter (2014, 4)


Digital Humanities and Computational Linguistics

Today

- ▶ Digital Humanities, broadly: Working with 'digital methods' on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area Reiter (2014, 4)
 - ▶ ... but this is a minority position in CL, often also seen as part of AI
 - ▶ Historically (and still today) split between engineering (natural language processing, NLP) and science/scholarship (computational linguistics, CL)
 - ▶  Neurolinguistic programming and natural language processing are **not the same** (both use 'NLP' as abbreviation)

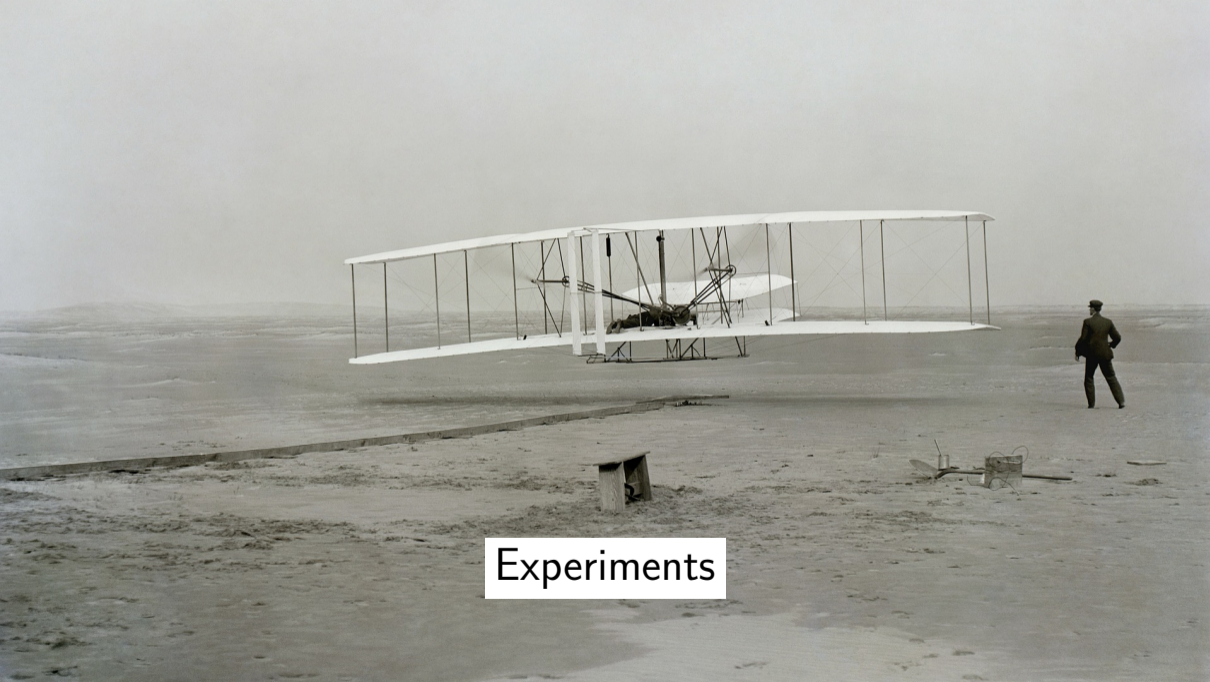
Digital Humanities and Computational Linguistics

Today

- ▶ Digital Humanities, broadly: Working with ‘digital methods’ on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area Reiter (2014, 4)
 - ▶ ... but this is a minority position in CL, often also seen as part of AI
 - ▶ Historically (and still today) split between engineering (natural language processing, NLP) and science/scholarship (computational linguistics, CL)
 - ▶  Neurolinguistic programming and natural language processing are **not the same** (both use ‘NLP’ as abbreviation)

University of Cologne

For historic reasons, CL and NLP are called “Sprachliche Informationsverarbeitung”



Experiments

Experiments

- ▶ Cornerstone of the 'scientific method'
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...

Experiments

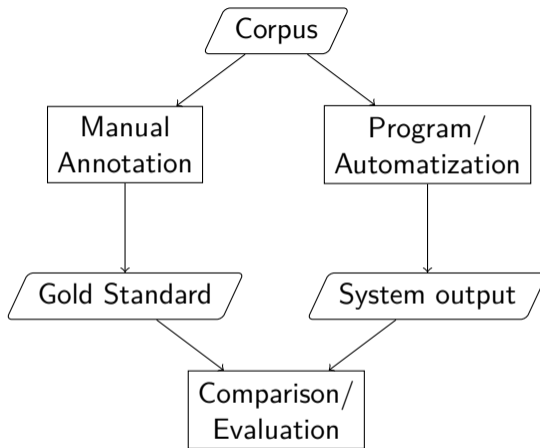
- ▶ Cornerstone of the 'scientific method'
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...
- ▶ Experiments are used to verify or falsify hypotheses
- ▶ Reproducibility: The outcome does not depend on the experimenter

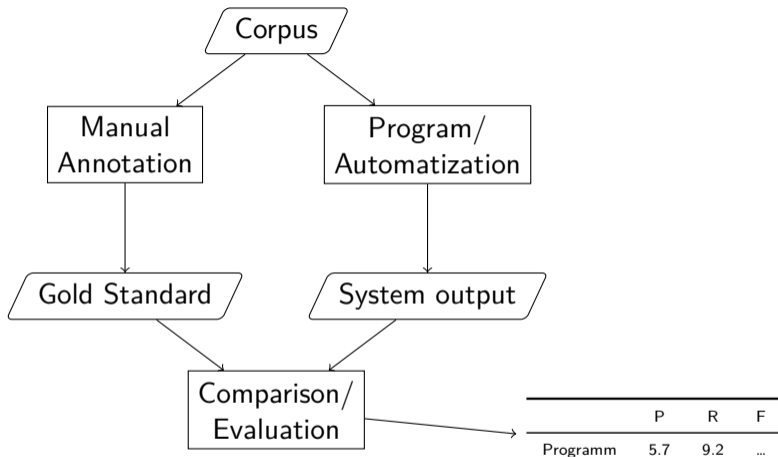
Experiments

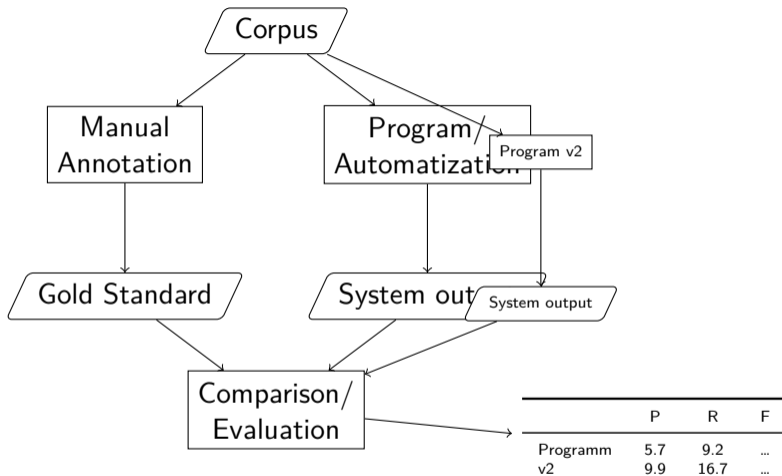
- ▶ Cornerstone of the 'scientific method'
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...
- ▶ Experiments are used to verify or falsify hypotheses
- ▶ Reproducibility: The outcome does not depend on the experimenter
- ▶ CL: Hypotheses about the operationalisation of language/text phenomena

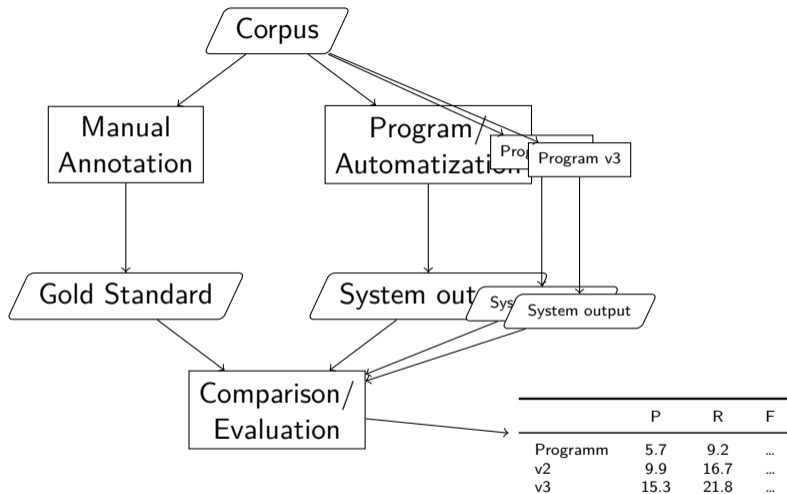
Example

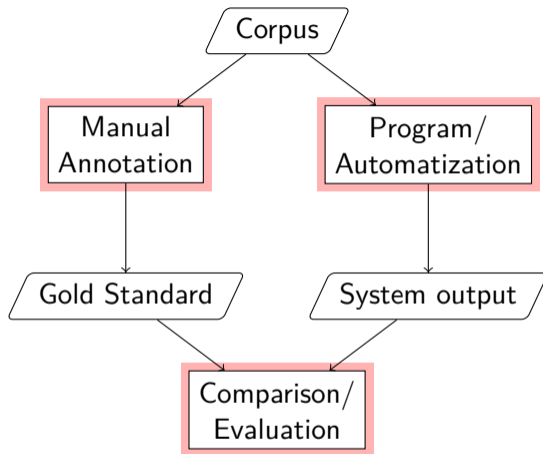
Position within a sentence is indicative for the part of speech











Manual Annotation

12 I have told you earthly things, and ye believe not, how shall ye believe, if I tell you of heavenly things?
13 And no man hath ascended up to heaven, but he that came down from heaven, even the Son of man which is in heaven.
14 And as Mo-ses lifted up the serpent in the wilderness, even so must the Son of man be lifted up:
15 That whosoever believeth in him should not perish, but have eternal life.
16 For God so loved the world, that he gave his only begotten Son, that whosoever believeth in him should not perish, but have everlasting life.
17 For God sent not his Son into the world to condemn the world; but that the world through him might be saved.
18 He that believeth on him is not condemned, but he that believeth not, is condemned already, because he hath not believed in the name of the only begotten Son of God.
19 And this is the condemnation, that light is come into the world, and men love darkness rather than light, because their deeds are evil.
20 For every one that doeth these things, hateth the light, and cometh not into the light, lest his deeds should be reproved.
21 But he that doeth the truth, cometh into the light, that his deeds may be manifested, which are done through truth.

And what he hath seen and heard, that he telleth, and no man receiveth his testimony. He saith that God is true.
35 The Father loveth the Son, and hath given all things into his hand.
36 He that believeth on the Son, and hath given life: and he that believeth not the Son, shall see life, but the wrath of God abideth on him.
The Woman at Jacob's Well
4 Phar-i-sees had heard that Je'-sus made and baptized more disciples than John, (Though Je'-sus himself baptized not, but his disciples.)
5 He left Ju-dae'-a, and departed again into Gal-i-lee.
6 And he must needs go through sea-side.
7 A certain Samaritan woman drew water into a well, and Je'-sus saith unto her, Give me to drink.
8 For I am thirsty.
9 She saith unto him, Sir, thou art a Jew, and I am a Samaritaness; how canst thou say unto me, Give me to drink? For the Jews have no fellowship with the Samaritans, because they hated them.
10 Je'-sus saith unto her, Whosoever drinketh of the water that I shall give him, shall never thirst; but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
11 She saith unto him, Sir, thou hast no bucket, and I have no way to draw water: for thou art a Jew.
12 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
13 She saith unto him, Sir, give me of this water.
14 Je'-sus saith unto her, If thou wilt, give me of this water.
15 She saith unto him, How canst thou, when thou hast no bucket, say unto me, Give me of this water?
16 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
17 She saith unto him, Sir, I see that thou art a prophet.
18 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
19 She saith unto him, Sir, I see that thou art a prophet.
20 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
21 She saith unto him, Sir, I see that thou art a prophet.
22 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
23 She saith unto him, Sir, I see that thou art a prophet.
24 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
25 She saith unto him, Sir, I see that thou art a prophet.
26 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
27 She saith unto him, Sir, I see that thou art a prophet.
28 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
29 She saith unto him, Sir, I see that thou art a prophet.
30 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
31 She saith unto him, Sir, I see that thou art a prophet.
32 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
33 She saith unto him, Sir, I see that thou art a prophet.
34 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
35 She saith unto him, Sir, I see that thou art a prophet.
36 Je'-sus saith unto her, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.

And upon this came his disciples, and said unto him, What speakest thou with her? Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
37 One of the Samaritan women saith unto her, I see that thou art a prophet.
38 Je'-sus answered and said, Yes, woman, thou art a Samaritaness, and thou sayest, I see that thou art a prophet.
39 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
40 She saith unto him, Sir, I see that thou art a prophet.
41 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
42 She saith unto him, Sir, I see that thou art a prophet.
43 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
44 She saith unto him, Sir, I see that thou art a prophet.
45 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
46 She saith unto him, Sir, I see that thou art a prophet.
47 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
48 She saith unto him, Sir, I see that thou art a prophet.
49 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
50 She saith unto him, Sir, I see that thou art a prophet.
51 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
52 She saith unto him, Sir, I see that thou art a prophet.
53 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
54 She saith unto him, Sir, I see that thou art a prophet.
55 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
56 She saith unto him, Sir, I see that thou art a prophet.
57 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
58 She saith unto him, Sir, I see that thou art a prophet.
59 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
60 She saith unto him, Sir, I see that thou art a prophet.
61 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
62 She saith unto him, Sir, I see that thou art a prophet.
63 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
64 She saith unto him, Sir, I see that thou art a prophet.
65 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
66 She saith unto him, Sir, I see that thou art a prophet.
67 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
68 She saith unto him, Sir, I see that thou art a prophet.
69 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
70 She saith unto him, Sir, I see that thou art a prophet.
71 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
72 She saith unto him, Sir, I see that thou art a prophet.
73 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
74 She saith unto him, Sir, I see that thou art a prophet.
75 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
76 She saith unto him, Sir, I see that thou art a prophet.
77 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
78 She saith unto him, Sir, I see that thou art a prophet.
79 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
80 She saith unto him, Sir, I see that thou art a prophet.
81 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
82 She saith unto him, Sir, I see that thou art a prophet.
83 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
84 She saith unto him, Sir, I see that thou art a prophet.
85 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
86 She saith unto him, Sir, I see that thou art a prophet.
87 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
88 She saith unto him, Sir, I see that thou art a prophet.
89 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
90 She saith unto him, Sir, I see that thou art a prophet.
91 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
92 She saith unto him, Sir, I see that thou art a prophet.
93 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
94 She saith unto him, Sir, I see that thou art a prophet.
95 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
96 She saith unto him, Sir, I see that thou art a prophet.
97 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
98 She saith unto him, Sir, I see that thou art a prophet.
99 Je'-sus answered and said, My bucket is here with me, but thou hast no water, because thou hast not drawn it: but the water that I shall give him, shall become in him a well of water, springing up to eternal life.
100 She saith unto him, Sir, I see that thou art a prophet.

Annotation

- ▶ Interdisciplinary 'false friend'
- ▶ Different meanings in different disciplines
 - ▶ Adding TEI/XML markup: DH community
 - ▶ Adding comments to page margins: Hermeneutic traditions
 - ▶ Literary studies, bible studies
 - ▶ Assigning categories to textual material: (computational) linguistics

Annotation

- ▶ Interdisciplinary 'false friend'
- ▶ Different meanings in different disciplines
 - ▶ Adding TEI/XML markup: DH community
 - ▶ Adding comments to page margins: Hermeneutic traditions
 - ▶ Literary studies, bible studies
 - ▶ Assigning categories to textual material: (computational) linguistics

Example

Der alte Mann wird verrückt .

Annotation

- ▶ Interdisciplinary 'false friend'
- ▶ Different meanings in different disciplines
 - ▶ Adding TEI/XML markup: DH community
 - ▶ Adding comments to page margins: Hermeneutic traditions
 - ▶ Literary studies, bible studies
 - ▶ Assigning categories to textual material: (computational) linguistics

Example

Der	alte	Mann	wird	verrückt	.
Artikel	Adjektiv	Nomen	Verb	?	Punkt

Annotation

- ▶ Interdisciplinary 'false friend'
- ▶ Different meanings in different disciplines
 - ▶ Adding TEI/XML markup: DH community
 - ▶ Adding comments to page margins: Hermeneutic traditions
 - ▶ Literary studies, bible studies
 - ▶ Assigning categories to textual material: (computational) linguistics

Example

Der	alte	Mann	wird	verrückt	.
Artikel	Adjektiv	Nomen	Verb	?	Punkt

Annotation Guidelines

- ▶ Describe the way to create the machine-readable truth
- ▶ What is to be annotated (which words)
- ▶ Working definitions or tests for categories
- ▶ Living documents: Need to be iteratively improved
- ▶ Community-wide accepted standards are needed

Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS

Anne Schüller, Simone Tenfel, Christine Stöckert
Universität Stuttgart
Institut für maschinelle Sprachverarbeitung

Christine Thielens
Universität Tübingen
Seminar für Sprachwissenschaft
[Draft]

14. November 1995

1.1 Zuweisung von Tags

Als allgemeine Regel gilt, daß jede Wortform genau ein Tag erfüllt. Der Begriff Wortform umfaßt neben "reinen" Wortformen auch Zahlen in Ziffern, Satzzeichen, Sonderzeichen (wie z.B. §, §), abgetrennte Wortteile oder Kompositions-Erstglieder (wie z.B. **Ein-** und **-Ausgang**) etc. Es wird davon ausgegangen, daß für das manuelle Taggen die Texte so aufbereitet sind, daß jede Zeile genau eine Wortform enthält.

1.2 Mehrwortlexeme

Damit ist es also (aus technischen Gründen) nicht möglich, Mehrwortlexeme als Ganzes zu taggen, oder kontraktive Formen mit einer Kombination aus mehreren Tags zu versehen. Kleinerweise sollen feststehende Ausdrücke wie *vor kurzem*, *vor allem* als Mehrwortlexeme (**multi word items**) aufgefaßt werden und von Tokenizer und Tagger so behandelt werden. Sollte dies technisch noch nicht möglich ist, werden als Kompromiß die einzelnen Teile annähernd so behandelt, als wenn die Teile einzeln stehen würden:

Beispiele:

- New/**NE** York/**NE** **nicht:** New York/**NE**
- so/**ADV** chab/**KOUI** **nicht:** so chab/**KOUI**
- zum/**APPART** **nicht:** zum/**APPART**

Bei aus 2 Teilen bestehenden Konjunktionen (*entweder - oder*, *entweder - noch*) werden **beide** Teile als **KON** getaggt. In den folgenden Guidelines werden Mehrwortlexeme durch das Zeichen **ml** gekennzeichnet, was besagt, daß diese Wortform idealerweise ein gemeinsames Tag bekommen sollte (welches hinter den Zeichen **ml** angegeben wird), als Kompromißlösung aber wie angegeben getaggt wird.

1.3 Behandlung von Abkürzungen

Es gibt kein eigenes Tag für Abkürzungen. Abgekürzte Wortformen werden generell so getaggt wie die ausgeschriebene Form. Abkürzungen für mehrere Wörter, die nicht durch Leerzeichen getrennt sind, werden entsprechend ihrer syntaktischen Funktion klassifiziert.

Beispiele:

- Herr/**NI** Dr./**NI** Meier/**NE**
- die/**genm**/**ADJA** Verhandlungen
- mit/**Haus** **tu**/**KIN** Garten
- z./**APPART** **B**/**NI**
- z.B./**ADV**
- d./**FGS** **tu**/**WFIN**
- chab/**KON**
- sondern/**KON**

Stuttgart-Tübingen Tagset (STTS)

- ▶ Welche Wortart hat das Wort 'verrückt'?

Stuttgart-Tübingen Tagset (STTS)

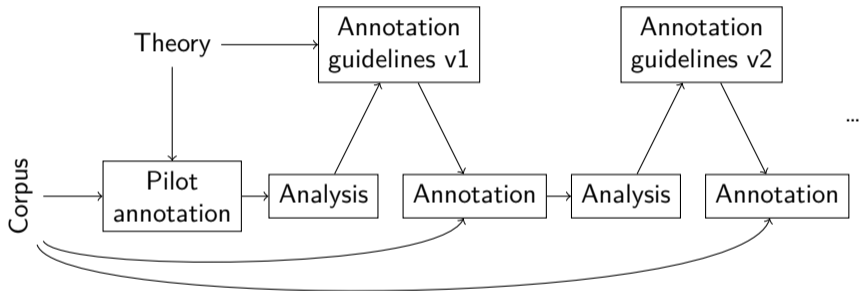
► Welche Wortart hat das Wort 'verrückt'?

Kriterien für Disambiguierung Kopulakonstruktionen mit ADJD vs. Verlaufspassiv mit VVPP:

- Verdacht auf VVPP: kann der Satz ins Aktiv gesetzt werden mit gleicher Semantik? Ja → VVPP
- von-PP oder ähnliche PP, die auf Verbsemantik hinweist → VVPP
- Ersetzung durch semantisch nahes Adjektiv möglich → ADJD

Abbildung: STTS Guidelines, S. 23

Annotation Workflow



Reiter2020ac; Hovy/Lavid (2010)

Annotation Analysis

- ▶ Multiple annotators annotate the same text(s)
- ▶ Annotations are compared
- ▶ Disagreements can be quantified ('Inter-Annotator-Agreement', IAA)

Cohen, 1960; Fleiss, 1971; Fournier, 2013; Mathet et al., 2015,

- ▶ Inter- und Intra-AA
- ▶ ...It's also a good idea to talk to the annotators

Indirect Annotations

- ▶ Annotations as a by-product of games
 - ▶ <https://www.artigo.org>
 - ▶ <https://anawiki.essex.ac.uk/phrasedetectives/>

Indirect Annotations

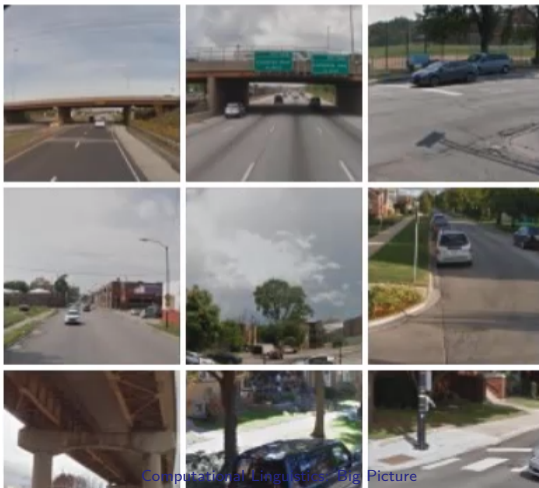
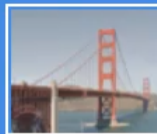
- ▶ Annotations as a by-product of games
 - ▶ <https://www.artigo.org>
 - ▶ <https://anawiki.essex.ac.uk/phrasedetectives/>
- ▶ Captchas for OCR correction



Indirect Annotation

- ▶ Annotations as a by-product
 - ▶ <https://www.robots.ox.ac.uk/~vgg/data/imagenet/>
 - ▶ <https://anaweb.org/>
- ▶ Captchas for OCR

Select all images with
bridges



Standard Corpora

- ▶ Iterative improvements of programs
- ▶ Measuring these improvements

- ▶ Accessibility
- ▶ Acceptance in the research community

Learning from Raw Data

Corpus annotation is expensive, therefore:

Learning from Raw Data

Corpus annotation is expensive, therefore:

- ▶ Train on things that are already there
- ▶ word2vec: Is 'dog' a context word of 'lazy'? Mikolov et al. (2013)
- ▶ BERT Devlin et al. (2019)
 - ▶ Can you fill in this blanked word? ("masked language modeling", MLM)
 - ▶ Are these two sentences natural neighbours? ("next sentence prediction", NSP)

Learning from Raw Data

Corpus annotation is expensive, therefore:





- ▶ Train on things that are already there
- ▶ word2vec: Is 'dog' a context word of 'lazy'? Mikolov et al. (2013)
- ▶ BERT Devlin et al. (2019)
 - ▶ Can you fill in this blanked word? ("masked language modeling", MLM)
 - ▶ Are these two sentences natural neighbours? ("next sentence prediction", NSP)
- ▶ Training data available in abundance
 - ▶ As long as there is digital data for a language
 - ⚠ Difficult to control what exactly is in there
 - ▶ More obvious for text-image data sets Birhane et al. (2021)

haveibeentrained.com



References II

-  Devlin, Jacob/Ming-Wei Chang/Kenton Lee/Kristina Toutanova (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: <https://aclanthology.org/N19-1423>.
-  Fleiss, Joseph L. (1971). “Measuring nominal scale agreement among many raters”. In: *Psychological Bulletin* 76.5, pp. 420–428.
-  Fournier, Chris (2013). “Evaluating Text Segmentation using Boundary Edit Distance”. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 1702–1712. URL: <http://aclweb.org/anthology/P13-1167>.

References III

-  Hovy, Eduard/Julia Lavid (2010). “Towards a ‘Science’ of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics”. In: *International Journal of Translation Studies* 22.1.
-  Lutz, Theo (1959). “Stochastische Texte”. In: *augenblick* 4, pp. 3–9. URL: https://www.netzliteratur.net/lutz%5C_schule.htm.
-  Mathet, Yann/Antoine Widlöcher/Jean-Philippe Métivier (2015). “The Unified and Holistic Method Gamma () for Inter-Annotator Agreement Measure and Alignment”. In: *Computational Linguistics* 41.3, pp. 437–479.
-  Mikolov, Tomáš/Kai Chen/Greg Corrado/Jeffrey Dean (2013). “Efficient Estimation of Word Representations in Vector Space”. In: *arXiv cs.CL*. URL: <https://arxiv.org/pdf/1301.3781.pdf>.

References IV

-  Nagao, Makoto (1984). “A Framework of a Mechanical Translation between Japanese and English by Analogy Principle”. In: *Proc. of the International NATO Symposium on Artificial and Human Intelligence*. Lyon, France: Elsevier North-Holland, Inc., pp. 173–180. ISBN: 0444865454.
-  Reiter, Nils (2014). “Discovering Structural Similarities in Narrative Texts using Event Alignment Algorithms”. PhD thesis. Heidelberg University, Germany.