

Komplexität und natürliche Sprache

Komplexität natürlicher Sprachen

Timm Lichte & Christian Wurm

Heinrich-Heine-Universität Düsseldorf

Sommersemester 2017, 21.07.2017



SFB 991



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

- 1 Rekapitulation
- 2 Vergleich natürlicher Sprachen

1 Rekapitulation

2 Vergleich natürlicher Sprachen

Konkrete Anwendungsfälle mit Bezug auf natürliche Sprache

- Klasse der natürlichen Sprachen: schwach kontextsensitiv, RTTM-erkennbar, ... (Sitzung am 03.05.)

Letzte Woche

- Vergleich natürlichsprachlicher Worte (=Sätze, Phrasen)
 - formale Komplexität
 - Quantitative Lesbarkeitsmetriken
 - (Phrasen-)Strukturelle Metriken

Heute

- Vergleich natürlicher Sprachen
 - Zum Beispiel: bilde die kürzesten TM's und vergleiche deren Länge
 - Problem: Durchführung, kulturtheoretische Implikationen

1 Rekapitulation

2 Vergleich natürlicher Sprachen

Gegeben eine natürliche Sprache NL :

- Funktion (Summe, Durchschnitt, ...) der Komplexität der Sätze in NL ?
 - Nein! Das kann nur bei endlichen Sprachen (oder Dokumenten) sinnvoll sein, und selbst da machen einzelne komplexe Sätze eine Sprache nicht notwendigerweise komplexer als eine Sprache aus einfachen Sätzen. Entscheidend ist die Varianz zwischen den Sätzen einer Sprache.
- Entropie von NL
 - Problem: Wahrscheinlichkeitsverteilung (egal welche) ist schwer abzuschätzen.
- Kolmogorov-Komplexität von NL
 - Problem: nicht berechenbar; eine Turingmaschine für NL aufzuschreiben, dürfte praktisch undurchführbar sein.

Grundproblem: *NL* ist in der Regel unendlich groß mit einem sehr großen Alphabet und vielen Konstruktionstypen. *NL* ist letztlich ein abstraktes Objekt.

Lösung: approximative (d.h. punktuelle) Komplexitätsbestimmung mittels

- Phoneminventar
- Silbeninventar
- Morpheminventar
- Konstruktionsinventar
- Teilsprachen

Fenk-Oczlon & Fenk [1]

Subsystems	Facets of linguistic complexity
phonology	size of phonemic inventory; syllable complexity (= n of phonemes/syllables); n of syllable types
morphology	complexity in word structure (n of morphemes and n of syllables per word); n of morphological cases, gender distinctions etc.; opaqueness of morphological forms
syntax	rigid (?) word order; hypotactic constructions
semantics	n of meanings per expression (homonymy, polysemy)

Fenk-Oczlon & Fenk [1]

- 1 The more syllables per word, the fewer phonemes per syllable.
- 2 The fewer phonemes per syllable, the more syllables per clause.
- 3 The more syllables per clause, the more syllables per word.
- 4 The more syllables per word, the fewer words per clause.

Kompensationshypothese

Komplexität in einem Bereich der Grammatik wird kompensiert durch die Einfachheit eines anderen Bereichs einer Grammatik.

“Equal-complexity dogma” (Parkvall 2008)

Alle Sprachen sind ungefähr gleich komplex.

Warum gibt es das “equal-complexity dogma”?

(Parkvall 2008)

- z.B. von Sapir
- gegen wissenschaftlichen Rassismus (“institutional racism”) Anfang des 20. Jh.
 - Komplexität ~ (semantische) Ausdrucksstärke ~ Wert

Warum gibt es das “equal-complexity dogma”?

(Parkvall 2008)

- z.B. von Sapir
- gegen wissenschaftlichen Rassismus (“institutional racism”) Anfang des 20. Jh.
 - Komplexität ~ (semantische) Ausdrucksstärke ~ Wert

Lösung: “Komplexität” und “Ausdrucksstärke” trennen!

- Pig Latin: Jede Sprache kann künstlich komplexer gemacht werden, ohne dass sie ausdrucksstärker wird.
- Esperanto: Sprachen ohne Irregularitäten sind deswegen nicht weniger ausdrucksstark.

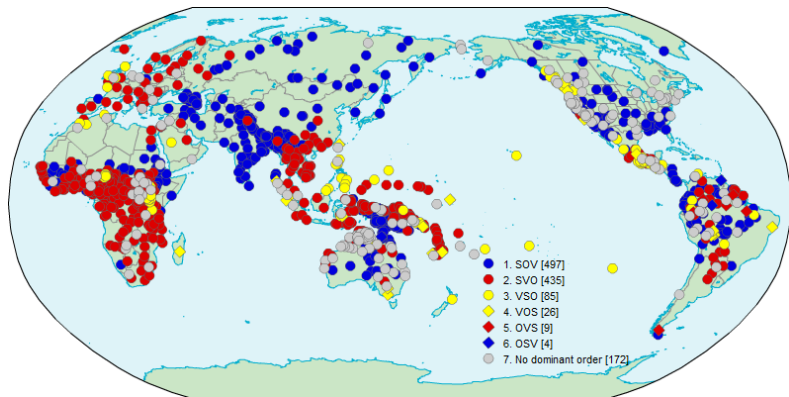
“all languages have the same expressive potential, while not necessarily being equally complex.”

Grundannahme: komplexere Sprache \Leftrightarrow komplexere Konstruktionen

Sprachvergleich mit WALS-Daten (Parkvall 2008)

Grundannahme: komplexere Sprache \Leftrightarrow komplexere Konstruktionen

Daten: Sprachmerkmale aus dem WALS (World Atlas of Linguistic Structures, 2005)



Merkmalsauswahl: Nicht alle Merkmale sind/scheinen relevant (z.B. Wortstellung, Synthese).

47 WALS-Merkmale + 6 eigene Merkmale

size of consonant inventories, size of vowel quality inventories, phonemic vowel nasalization, complexity of syllable structure, tone, overt marking of direct object, double marking of direct object, possession by double marking, overt possession marking, reduplication, gender, number of genders, non-semantic gender assignment, grammaticalized nominal plural, definite articles, indefinite articles, ...

Merkmalsauswahl: Nicht alle Merkmale sind/scheinen relevant (z.B. Wortstellung, Synthese).

47 WALS-Merkmale + 6 eigene Merkmale

size of consonant inventories, size of vowel quality inventories, phonemic vowel nasalization, complexity of syllable structure, tone, overt marking of direct object, double marking of direct object, possession by double marking, overt possession marking, reduplication, gender, number of genders, non-semantic gender assignment, grammaticalized nominal plural, definite articles, indefinite articles, ...

Merkmalsinterpretation: Werte werden in den Zahlenraum zwischen 0 und 1 abgebildet.

Merkmalshäufigkeit: nur Sprachen mit mindestens 30 von 53 Merkmalen

Metrik je Sprache: $(F_1 + \dots + F_{53}) / |\text{spezifizierte Merkmale}|$

- Die Indoeuropäischen Sprachen befinden sich eher am oberen Ende (“Eurocentrism ?”).
- davon abgesehen: kompatibel mit den Erwartungen von Typologen
- Die einzig vertretenen Kreolsprachen, Ndyuka and Sango, sind auf Position 150 und 155 (von 155 Sprachen).

- Die Indoeuropäischen Sprachen befinden sich eher am oberen Ende (“Eurocentrism ?”).
- davon abgesehen: kompatibel mit den Erwartungen von Typologen
- Die einzig vertretenen Kreolsprachen, Ndyuka and Sango, sind auf Position 150 und 155 (von 155 Sprachen).

Zwei Reaktionen:

- Kreolsprachen haben vergleichsweise einfache Grammatiken (McWhorter)
- Solche Befunde sind “Mythen”, “Vorurteile”, “Illusionen”, “Fantasien” (DeGraff)

Parkvalls Schlussfolgerung:

strukturelle Komplexität \neq psycholing. Komplexität \neq (sem.) Ausdrucksstärke

“complexity of a language correlates with its age” (McWhorter)

Wenn sich Sprachen in ihrer Komplexität unterscheiden, tun das **nicht** auch zwangsläufig die dranhängenden Kulturen!

Falsche Anwendung der Sapir-Whorf-Hypothese.

Sapir-Whorf-Hypothese

Sprache \Leftrightarrow Denken (\Leftrightarrow Kultur)

- [1] Fenk-Oczlon, Gertraud & August Fenk. 2008. Complexity trade-offs between the subsystems of language. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change* (Studies in Language Companion Series 94), 43–65. Amsterdam: John Benjamins.
- [2] Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. (Studies in Language Companion Series 94). Amsterdam: John Benjamins.
- [3] Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change* (Studies in Language Companion Series 94), 265–285. Amsterdam: John Benjamins.