Complexity in grammar

Kritik des algorithmischen Komplexitätsbegriffs
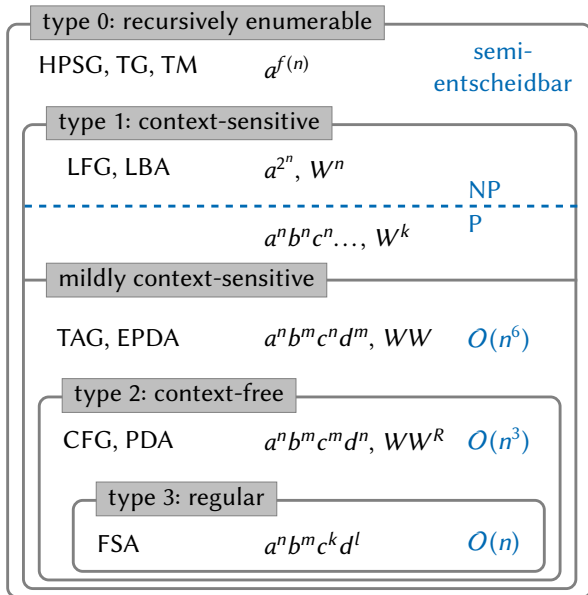
Timm Lichte

HHU Düsseldorf

WS 2015/2016, 04.11.2015

Was verursacht Komplexität?

- Größe (von Alphabet/Wort/Grammatik)
- Entropie (eines Wortes)
  - ⇒ informatische Komplexität
- Beschreibungsaufwand (für Alphabet/Wort/Sprache/Grammatik)
  - ⇒ Beschreibungskomplexität
- Mächtigkeit (einer Sprache/Grammatik)
  - ⇒ relative extensionale Komplexität
- Berechnungsaufwand (des Wortproblems einer Sprache)
  - ⇒ algorithmische Komplexität

Letzte Sitzung: algorithmische × extensionale Komplexität

type 0: recursively enumerable

HPSG, TG, TM　　　$a^{f(n)}$　　　semi-entscheidbar

type 1: context-sensitive

LFG, LBA　　　a^{2^n}, W^n

NP

P

$a^n b^n c^n \ldots$, W^k

mildly context-sensitive

TAG, EPDA　　　$a^n b^m c^n d^m$, WW　　$O(n^6)$

type 2: context-free

CFG, PDA　　　$a^n b^m c^m d^n$, WW^R　　$O(n^3)$

type 3: regular

FSA　　　$a^n b^m c^k d^l$　　$O(n)$

also nur schwach kontextsensitive Grammatikformalismen???

Zwei kritische Stimmen

Berwick, Robert C. & Amy S. Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

- Verteidigung der Transformationsgrammatik/GB-Theorie
- ⇒ Grammatikgröße hat einen wesentlichen Einfluss!

Pollard, Carl. 1996. The nature of constraint grammar. Paper presented at the 11th Pacific Conference of Language, Information and Computation (PACLIC).

- Verteidigung der HPSG
- ⇒ Natürliche Sprache ist womöglich semi-entscheidbar!
- ⇒ Theorien sollte sich aller Mittel bedienen dürfen!

Berwick & Weinberg (1984): Aufbau

Kap. 1 "Computational and linguistic theory"
- Grundbegriffe: TM, Komplexität, TG, GB

Kap. 2 "Grammars and models of language use"
- Competenz vs. Performanz, Derivational Theory of Complexity

Kap. 3 "Generative power and linguistic theories"
- Grammatikgröße, effizientes Parsen, asymptotische Komplexität

Kap. 4 "Transformational Grammar: A formal analysis"
- Kritik an Lexical-Functional Grammar (LFG)

Kap. 5 "Parsing and Government-Binding Theory"
- Marcus Parser

Kap. 6 "Language acquisition and Transformational Grammar: A modern view"

Chomsky-Hierarchie versus algorithmische Komplexitätsklassen:

> What is the relationship between the Chomsky language classes and time complexity classes? As it turns out, **there is no clean one-to-one correspondence between them**. This lack of fit is a problem for anyone who would like to argue on the basis of a time complexity result that one or another Chomsky language class is more "natural" than another; (S.9)

> Context-sensitive languages take linear space for their recognition by definition and at worst exponential time. But some strictly context-sensitive languages, languages that are not context-free, are recognizable in real time; for example, the "triple counting" language $a^n b^n c^n$ is so recognizable. (S.10)

Erkennung versus Parsing:

> Yet another stumbling block is the difference between just recognizing a language and parsing it. Recognizing a language is easier. **All the TM has to do is accept or reject sentences. Parsing is at least this hard.** Not only does the machine have to say whether or not a sentence is in a given language, but also it must write down a representation of how that sentence was derived with respect to some fixed grammar. […] Parsing, then, is grammar relative; recognition is not. (S.10)

> But are such languages always parsable in real time, given a grammar? The answer is no. (S.10)

Beispiel: $ca^+(d|e)$
CFG: $S \rightarrow cX$, $S \rightarrow cY$, $X \rightarrow aX$, $X \rightarrow aD$,
$\qquad Y \rightarrow aY$, $Y \rightarrow aE$, $E \rightarrow e$, $D \rightarrow d$

Vorteile von "efficient recognizability":

- psychological plausability; grammar can be transparently embedded
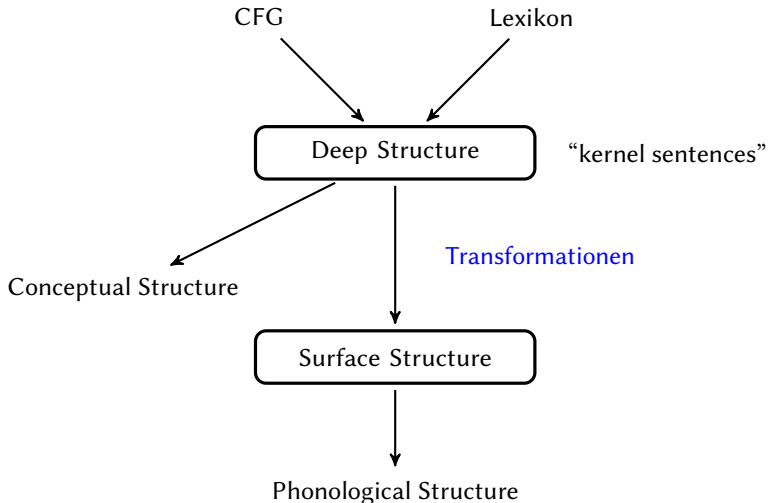- learnability of a language

Aber: Sprachen der Transformationsgrammatik können semi-entscheidbar sein.
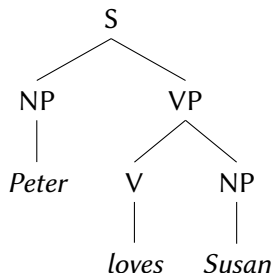
Ziel: Rehabilitation von TG

Exkurs: Transformationsgrammatik
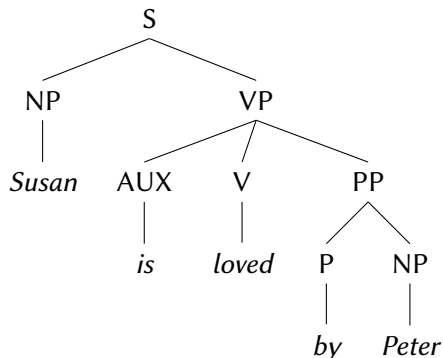
Chomsky (1957; 1965)

CFG   Lexikon

Deep Structure   "kernel sentences"

Conceptual Structure

Transformationen

Surface Structure

Phonological Structure

Exkurs: Transformationsgrammatik

Chomsky (1957; 1965)

"kernel sentences"

$$NP_1 \ V \ NP_2 \rightarrow NP_2 \ AUX \ V \ by \ NP_1$$

- Komplexitätstheoretisch sehr problematisch: Das Wortproblem für TG ist **unentscheidbar** (Peters & Ritchie 1973).

Berwick & Weinberg (1984): Kapitel 3

Rehabilitation von TG: Strategien

- Transformationen entfernen: GPSG (Gazdar 1981), LFG (Kaplan & Bresnan 1982)
- Transformationen einschränken: Extended Standard Theory (decidable), GB
- Zeigen, dass natürliche Sprache ineffizient/semi-entscheidbar ist (Hintikka 1974; Chomsky 1980)
- **Konkurrenten schwächen: GPSG/CFG & LFG**
- **algorithmische Komplexität kritisieren**

Gazdar's argument (S.87)

(1) People parse sentences rapidly.

(2) The sentences of any context-free language can be parsed rapidly.

(3) Gazdar's phrase-structure grammars generate only context-free languages.

(4) Not all languages generated by transformational grammars can be parsed rapidly.

- Conclusion: The theory of transformational grammar cannot provide an explanation of how people parse sentences; in contrast, a theory that can weakly generate only context-free languages can provide such an explanation.

"But is this line of argument valid? We believe not."

$$NL \neq CFL \subset P$$

> We shall see that biologically relevant parsing efficiency need not be primarily determined by general, mathematically defined measures of efficiency. Therefore, although mathematical efficiency measures may apply in an abstract, formal sense to rank context-free phrase structure grammars as "better" than transformational grammars, this ranking may be **relevant only in formal theory, not in biological practice**. (S.88)

"biological practice" = Parsen von konkreten Sätzen mit konkreten Grammatiken

⇒ Verhalten bei realistischen Satzlängen ("relevant range problem")
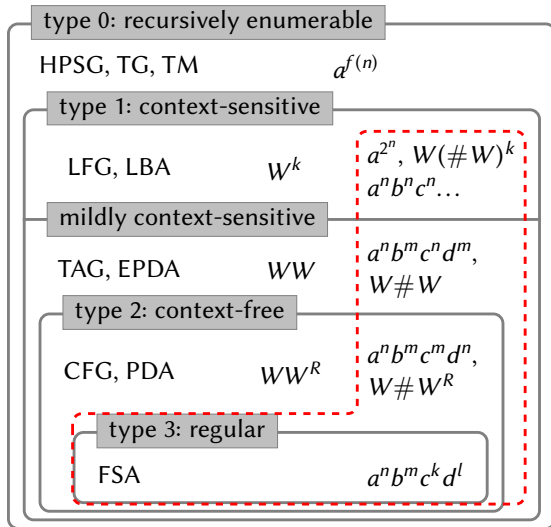
⇒ Rolle der Grammatikgröße ("implementation problem")

Was ist "efficient parsability"? (S.89–93)

- kubisch / $O(n^3)$
- PTIME
- "real time"

⇒ PTIME und "real time" entsprechen keinem Chomsky-Typ!

type 0: recursively enumerable

HPSG, TG, TM $\quad a^{f(n)}$

type 1: context-sensitive

LFG, LBA $\quad W^k \qquad a^{2^n}, W(\#W)^k$
$a^n b^n c^n \ldots$

mildly context-sensitive

TAG, EPDA $\quad WW \qquad a^n b^m c^n d^m,$
$W \# W$

type 2: context-free

CFG, PDA $\quad WW^R \qquad a^n b^m c^m d^n,$
$W \# W^R$

type 3: regular

FSA $\qquad a^n b^m c^k d^l$

deterministic
real-time
definable
(Rosenberg 1967)

"asymptotic complexity"

- kn^2 versus $k'n^3$: Je größer n, desto unwichtiger k und k'.

$$\lim_{n \to \infty} kn^2 < k'n^3$$

- k und k' repräsentieren "implementation details", von denen abstrahiert wird.
- Frage: Ist das auch "biologically relevant"?

In contrast cognitive measures must focus on what range of input sentence lengths are actually encountered in biological practice, for the "constant" values in front of the functional forms are proxies for the mental representations that parsing algorithms presumably are to use.
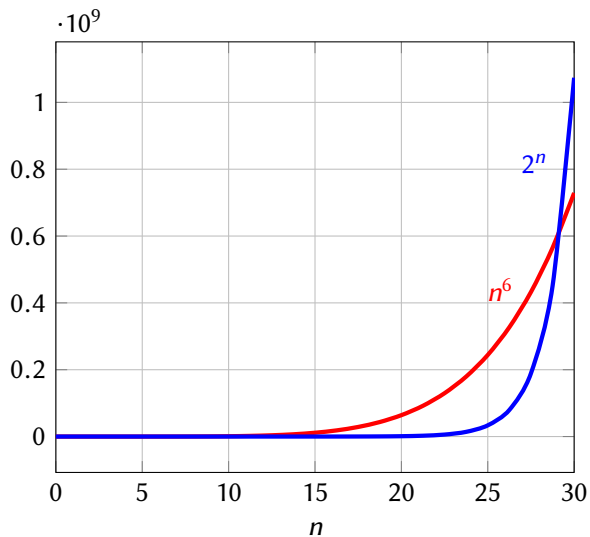
- "relevant range problem" $\Rightarrow$ realistischer Bereich für n
- "implementation problem" $\Rightarrow$ Einfluss von k und k' in diesem Bereich

"relevant range" bei n^3 versus 2^n
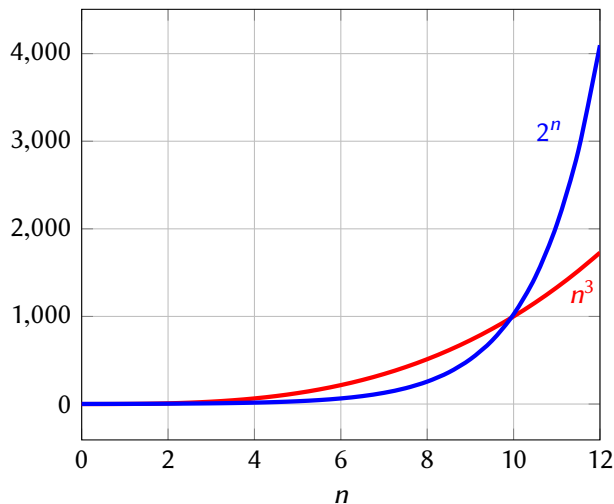
"relevant range" bei n^6 versus 2^n

"implementation" = Grammatikgröße
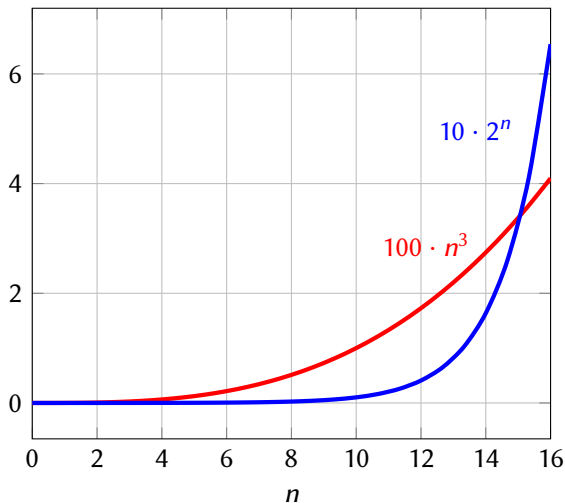
Beispiel für den Einfluss der Grammatikgröße

Komplexität des Early-Algorithmus:[5] $k|G|^2 n^3$

Sei $|G| = 500$ und $n = 20$:

$$\frac{500^2}{20^3} \approx \frac{30}{1}$$

$$500^2 \approx 63^2$$

Einfluss der "implementation" bei n^3 versus 2^n

Nota bene:
$|G|$ kann sich erheblich (z.B. exponentiell) unterscheiden!

Berwick & Weinberg (1984): Zusammenfassung

- Die Chomsky-Hierarchie hat keinen passenden Typ für natürliche Sprache.
- Asymptotische algorithmische Komplexität ist zu abstrakt.
- Tatsächliche algorithmische Komplexität kann bei mächtigeren Grammatiken geringer sein.
- ⇒ Mächtige Grammatikformalismen sind unsere Freunde!

Pollard (1996)

Ist die Mächtigkeit/Komplexität eines Grammatikformalismus wirklich so wichtig?

Pollard (1996)

If physicists required the formalism to constrain the theory

Editor: *Professor Einstein, I'm afraid we can't accept this manuscript of yours on general relativity.*

Einstein: *Why? Are the equations wrong?*

Editor: *No, but we noticed that your differential equations are expressed in the first-order language of set theory. This is a totally unconstrained formalism! Why, you could have written down ANY set of differential equations!*

⇒ Der Formalismus ist nicht Teil der Theorie! Deshalb sollte man den Formalismus so mächtig/bequem wir möglich machen. (Bei Pollard: HPSG)

Pollard (1996)

Sollte man die Mächtigkeit berücksichtigen? Zwei Standpunkte:

Nein! HPSG/Pollard (1996): Der Formalismus ist nicht Teil der Theorie! Deshalb sollte man den Formalismus so mächtig/bequem wir möglich machen.

Ja! TAG: Der Formalismus ist Teil der Theory, deshalb soll er so komfortabel wie möglich und so mächtig wie nötig sein!

Geschmacksfrage? Letztendlich ja!

Richard Feynman 1964 (The Character of Physical Law)   (LINK)

Suppose you have two theories A and B, which look completely different psychologically with different ideas in them. But all the consequences that are computed are exactly the same. Suppose we have two such theories, how are we gonna decide which one is right: no way, not by science!

However, for psychological reasons, in order to guess new theories, these two theories are very far from equivalent. Therefore we must keep all the theories in our head.

[1]   Berwick, Robert C. & Amy S. Weinberg. 1984. *The grammatical basis of linguistic performance: Language use and acquisition.* Cambridge, MA: MIT Press.

[2]   Chomsky, Noam. 1957. *Syntactic structures.* Den Haag: Mouton.

[3]   Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

[4]   Chomsky, Noam. 1980. *Rules and representations.* Oxford, UK: Basil Blackwell.

[5]   Earley, Jay. 1970. An efficient context-free parsing algorithm. *Commun. ACM* 13(2). 94–102. http://doi.acm.org/10.1145/362007.362035.

[6]   Gazdar, Gerald. 1981. Unbounded dependencies and coordinated structure. *Linguistic Inquiry* 12. 155–182.

[7]   Hintikka, Jaakko. 1974. Quantifiers vs. quantification theory. *Linguistic Inquiry* 5(2). 153–177. http://www.jstor.org/stable/4177815.

[8]   Kaplan, Ron M. & Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: MIT Press.

[9]   Peters, Stanley P. & Robert W. Ritchie. 1973. On the generative power of transformational grammar. *Information Sciences* 6. 49–83.

[10]   Pollard, Carl. 1996. The nature of constraint grammar. Paper presented at the 11th Pacific Conference of Language, Information and Computation (PACLIC).

[11]   Rosenberg, Arnold L. 1967. Real-time definable languages. *Journal of the Association for Computing Machinery* 14(4). 645–662.