English Syntax and Context Free Grammars

COMP-599

Oct 5, 2016

Gradient Descent Summary

Descent vs ascent

Convention: think about the problem as a minimization problem

Minimize the *negative* log likelihood

Initialize $\theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$ randomly

Do for a while:

Compute $\nabla l(\theta)$, which will require dynamic programming (i.e., forward algorithm)

$\theta \leftarrow \theta - \gamma \nabla l(\theta)$

Stochastic Gradient Descent

In the standard version of the algorithm, the gradient is computed over the entire training corpus.

- Weight update only once per iteration through training corpus.

Alternative: calculate gradient over a small mini-batch of the training corpus and update weights then

- Many weight updates per iteration through training corpus
- Usually results in much faster convergence to final solution, without loss in performance

Stochastic Gradient Descent

Initialize $\theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$ randomly

Do for a while:

Randomize order of samples in training corpus

For each mini-batch in the training corpus:

Compute $\nabla^* l(\theta)$ over this mini-batch

$\theta \leftarrow \theta - \gamma \nabla^* l(\theta)$

Outline

What is Syntax

English Syntax

Context Free Grammars

Syntax

How words can be arranged together to form a **grammatical** sentence.

- *This is a valid sentence.*
- **A sentence this valid is.*

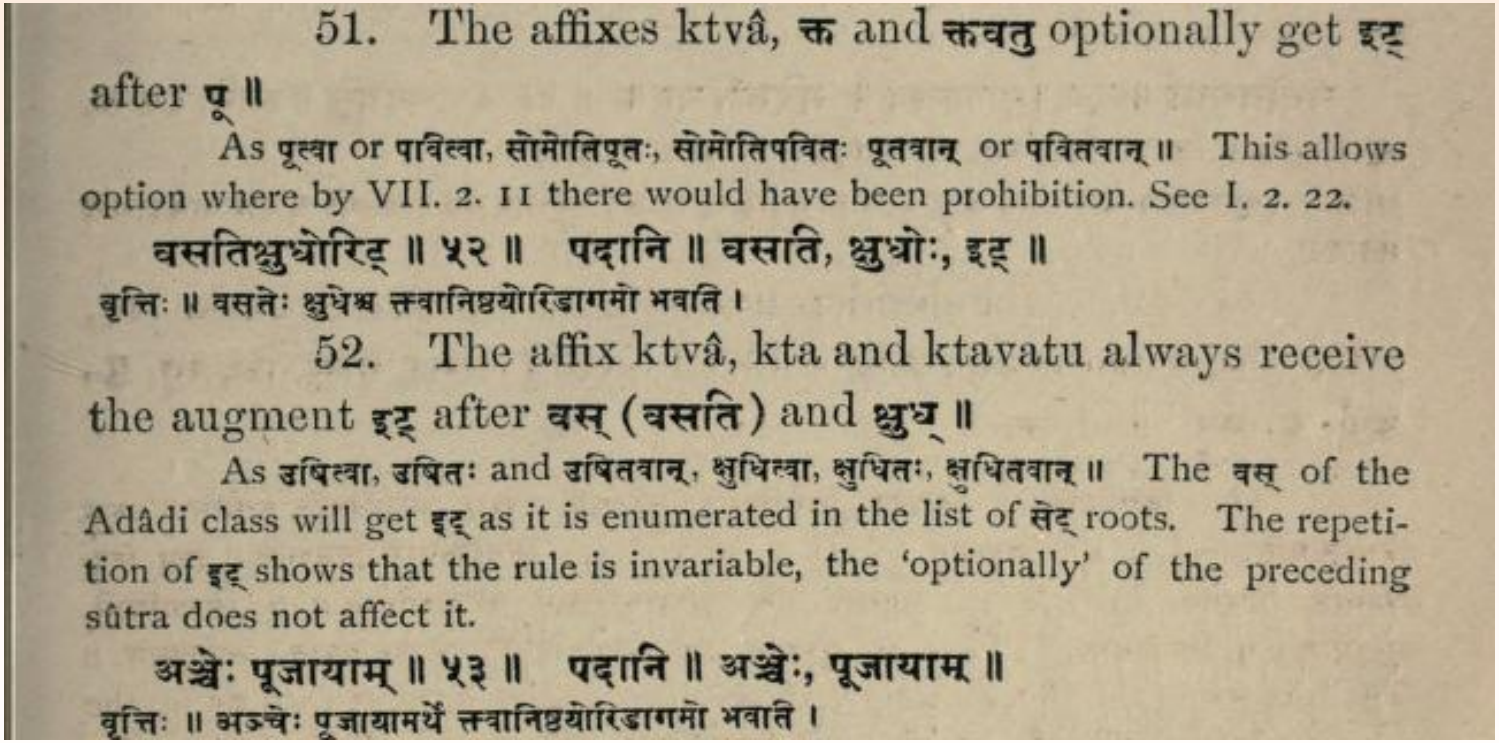
An asterisk is used to indicate **ungrammaticality**.

One view of syntax:

> Generate all and exactly those sentences of a language which are grammatical

The First Grammarian

Panini (Pāṇini) from the 4th century B.C. developed a grammar for Sanskrit.



51. The affixes ktvâ, क्त and क्तवतु optionally get इट् after पू ॥

As पूत्वा or पवित्वा, सोमोतिपूतः, सोमोतिपवितः पूतवान् or पवितवान् ॥ This allows option where by VII. 2. 11 there would have been prohibition. See I. 2. 22.

वसतिक्षुधोरिट् ॥ ५२ ॥ पदानि ॥ वसति, क्षुधोः, इट् ॥

वृत्तिः ॥ वसतेः क्षुधेश्च क्त्वानिष्ठयोरिडागमो भवति ।

52. The affix ktvâ, kta and ktavatu always receive the augment इट् after वस् (वसति) and क्षुध् ॥

As उषित्वा, उषितः and उषितवान्, क्षुधित्वा, क्षुधितः, क्षुधितवान् ॥ The वस् of the Adâdi class will get इट् as it is enumerated in the list of सेट् roots. The repetition of इट् shows that the rule is invariable, the 'optionally' of the preceding sûtra does not affect it.

अञ्चेः पूजायाम् ॥ ५३ ॥ पदानि ॥ अञ्चेः, पूजायाम् ॥

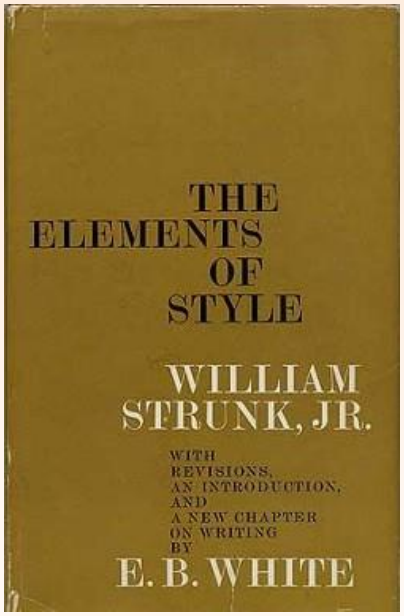
वृत्तिः ॥ अञ्चेः पूजायामर्थे क्त्वानिष्ठयोरिडागमो भवति ।

Source: https://archive.org/details/ashtadhyayitrans06paniuoft

What We Don't Mean by Grammar

Rules or guides for how to write properly

e.g.,



These style guides are **prescriptive**. We are concerned with **descriptive** grammars of naturally occurring language.

Basic Definitions Terms

- grammaticality
- prescriptivism vs descriptivism
- constituency
- grammatical relations
- subcategorization

Constituency

A group of words that behave as a unit

Noun phrases:

- *computational linguistics*, *it*, *Justin Trudeau*, *three people on the bus*, *"Jean-Claude Van Damme, the Muscles from Brussels"*

Adjective phrases:

- *blue*, *purple*, *very good*, *ridiculously annoying and tame*

Tests for Constituency

1. They can appear in similar syntactic environments.

I saw …

it

Jean-Claude Van Damme, the Muscles from Brussels

three people on the bus

**Van*

**on the*

Tests for Constituency

2. They can be placed in different positions or replaced in a sentence as a unit.

[Jean-Claude Van Damme, the Muscles from Brussels], beat me up.

It was [Jean-Claude Van Damme, the Muscles from Brussels], who beat me up.

I was beaten up by [Jean-Claude Van Damme, the Muscles from Brussels].

He beat me up. (i.e., J-C V D, the M from B)

Tests for Constituency

3. It can be used to answer a question.

Who beat you up?

[Jean-Claude Van Damme, the Muscles from Brussels]

**[the Muscles from]*

Grammatical Relations

Relationships between different constituents

Subject

- *Jean-Claude Van Damme relaxed.*
- *The wallet was stolen by a thief.*

(Direct) object

- *The boy kicked the ball.*

Indirect object

- *She gave him a good beating.*

There are many other grammatical relations.

Subcategorization

Notice that different verbs seem to require a different number of **arguments**:

relax	1	subj
steal*	2	subj, dobj
kick	2	subj, dobj
give	3	subj, iobj, dobj

*the passive changes the subcategorization of the verb

More Subcategorization

Some other possibilities:

want 2 subj, inf. clause

- *I want to learn about computational linguistics.*

apprise 3 subj, obj, pobj with of

- *The minister apprised him of the new developments.*

different 2 subj, pobj with from/than/to

- *This course is different [from/than/to] what I expected.*

Short Exercise

Identify the prepositional phrase in the following sentence. Give arguments for why it is a constituent.

The next assignment is due on Wednesday, October 19th.

Formal Grammars

Since we are computational linguists, we will use a formal computational model of grammar to account for these and other syntactic concerns.

Formal grammar

Rules that generate a set of strings that make up a **language**.

(In this context, language simply refers to a set of strings.)

Why?

- Formal understanding lets us develop appropriate algorithms for dealing with syntax.
- Implications for cognitive science/language learning

FSAs and Regular Grammars

We've already seen examples of languages defined by formal grammars before this class!

FSAs to describe aspects of English morphology

- An FSA generates a **regular language**
- FSAs correspond to a class of formal grammars called **regular grammars**

To describe the syntax of natural languages (with multiple constituents, subcategorization, etc.), we need a more powerful class of formal grammars – **context free grammars** (**CFG**s).

Context Free Grammars (CFG)s

Rules that describe what possible sentences are:

S → NP VP

NP → this

VP → V

V → is | kicks| jumps | rocks

Constituent Tree

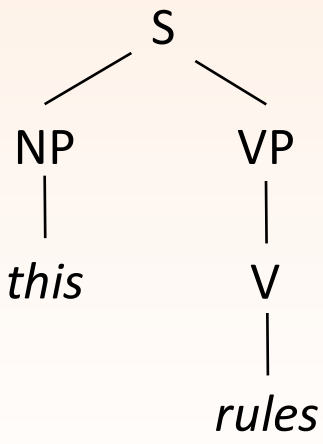
Trees (and sentences) generated by the previous rules:

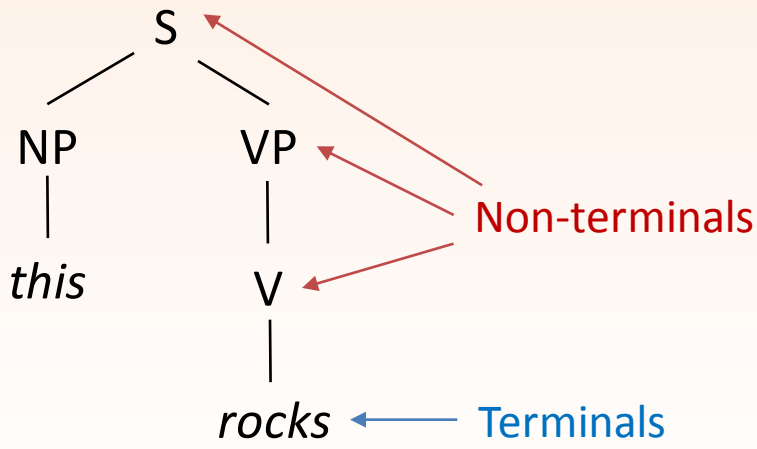
S → NP VP

VP → V

NP → *this*

V → *is* | *rules*| *jumps* | *rocks*





Formal Definition of a CFG

A 4-tuple:

- N set of **non-terminal** symbols
- Σ set of **terminal** symbols
- R set of **rules** or **productions** in the form $A \rightarrow (\Sigma \cup N)^*$, and $A \in N$
- S a designated **start symbol**, $S \in N$

Extended Example

Let's develop a CFG that can account for verbs with different subcategorization frames:

intransitive verbs	*relax*	1	subj
transitive verbs	*steal, kick*	2	subj, dobj
ditransitive verbs	*give*	3	subj, iobj, dobj

Undergeneration and Overgeneration

Problems with above grammar:

Undergeneration: misses valid English sentences

- *The boy kicked the ball softly*.
- *The thief stole the wallet with ease.*

Overgeneration: generates ungrammatical sentences

- **The boy kick the ball.*
- **The thieves steals the wallets.*

Extension 1

Let’s add adverbs and prepositional phrases to our grammar

Recursion

Consider the following sentences:

- *The dog barked.*
- *I know that the dog barked.*
- *You know that I know that the dog barked.*
- *He knows that you know that I know that the dog barked.*
- …

In general:

S -> NP VP

VP -> Vintr

Vintr -> *barked*

VP -> Vthat Sthat

Vthat-> *know*

Sthat -> *that* **S**

Recursion

This recursion in the syntax of English means that sentences can be infinitely long (theoretically).

- For a given sentence S, you can always make it longer by adding [I/you/he know(s) that S].

In practice, the length is limited because we have limited attention span/memory/processing power.

Exercise

Let's try to fix the subject-verb agreement issue:

Present tense:

- Singular third-person subject -> verb has affix of –s or –es
- Otherwise -> base form of verb
- (*to be* is an exception, along with other irregular verbs)

Dependency Grammar

Grammatical relations induce a **dependency relation** between the words that are involved.

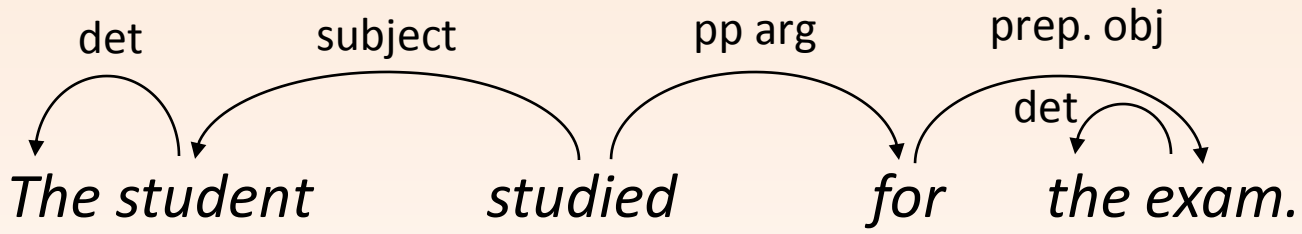
The student studied for the exam.

Each phrase has a **head word**.

- the student studied for the exam
- the student
- for the exam
- the exam

Dependency Grammar

We can represent the grammatical relations between phrases as directed edges between their heads.



This lets us get at the relationships between words and phrases in the sentence more easily.

Who/what are involved in the studying event?

- student, for the exam

Converting between Formalisms

Dependency trees can be converted into a standard constituent tree deterministically (if the dependency edges don't cross each other).

Constituent trees can be converted into a dependency tree, if you know what is the **head** of the constituent.

Let's convert some of our previous examples…

Crossing Dependencies

Yes, there can be crossing dependencies.

Especially if the language has **freer word order**:

Er hat mich versucht zu erreichen.

Er hat versucht mich zu erreichen.

He tried to reach me.

These have the same literal meaning.

Crossing Dependencies Example

What would the dependency edges be in these cases?

Er hat versucht, mich zu erreichen.

HE HAS TRIED ME TO REACH

Er hat mich versucht zu erreichen.

HE HAS ME TRIED TO REACH

Notice the discontinuous constituent that results in the second case.