Doktorarbeitsvorhaben

How to deal with grammar

Yvonne Meeres

Contents

- Motivation, Idea and Outcome
- Signal Processing
- Robustness
- Generative Grammar
- Summary of the whole job ...

The idea:
Why is natural language so difficult and easy?

Outcome

Hypothesis: Optimal Communication System

Every natural language is an optimal »technical« communication system for human communication.

From this hypothesis follows:

- Derivation of algorithms for technical robust communication
- Precise tool suite for the analysis of language learning

Trivial if formalized

After formalizing some insights are trivial.

Phone and Phoneme

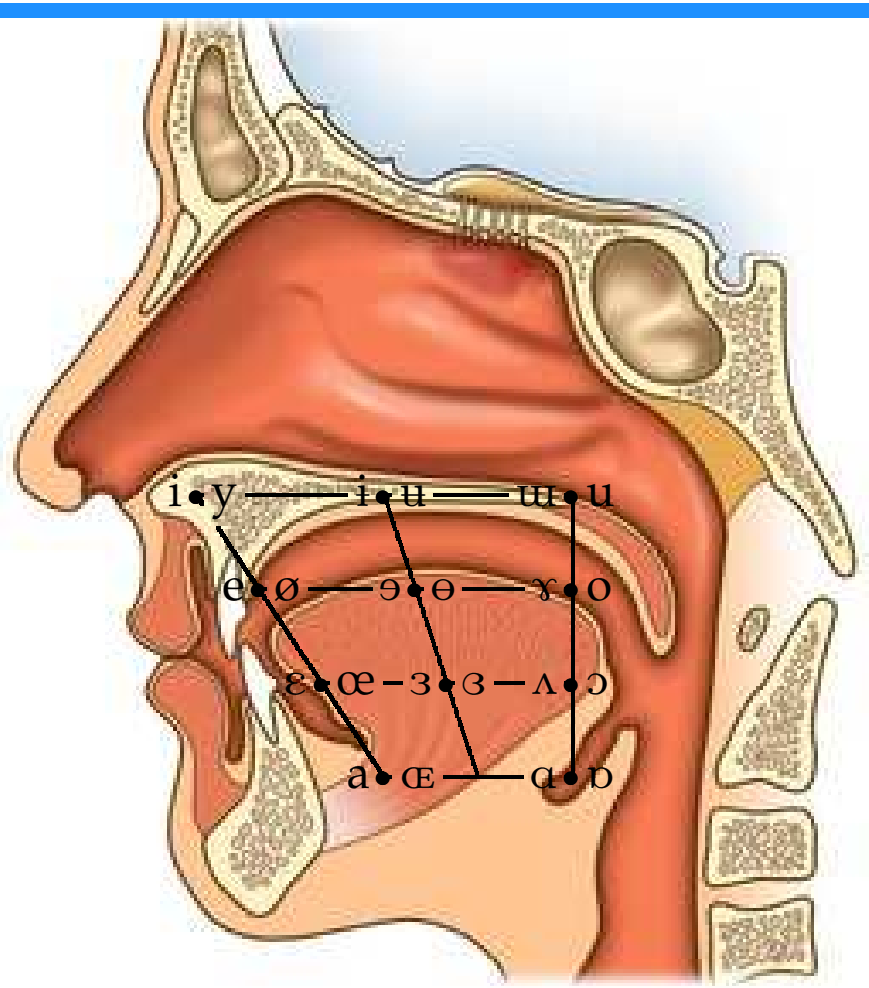
DEFINITION: Phone– and Phoneme–Vector

A ***Phone-Vector*** or a ***Phoneme-Vector*** is a vector which components $a_i \in \mathbb{R}$ represent the articulatory characteristics of a phoneme or a phone, respectively.
For Phones a_i is $\in \mathbb{R}$, for Phonemes $\in \mathbb{N}$

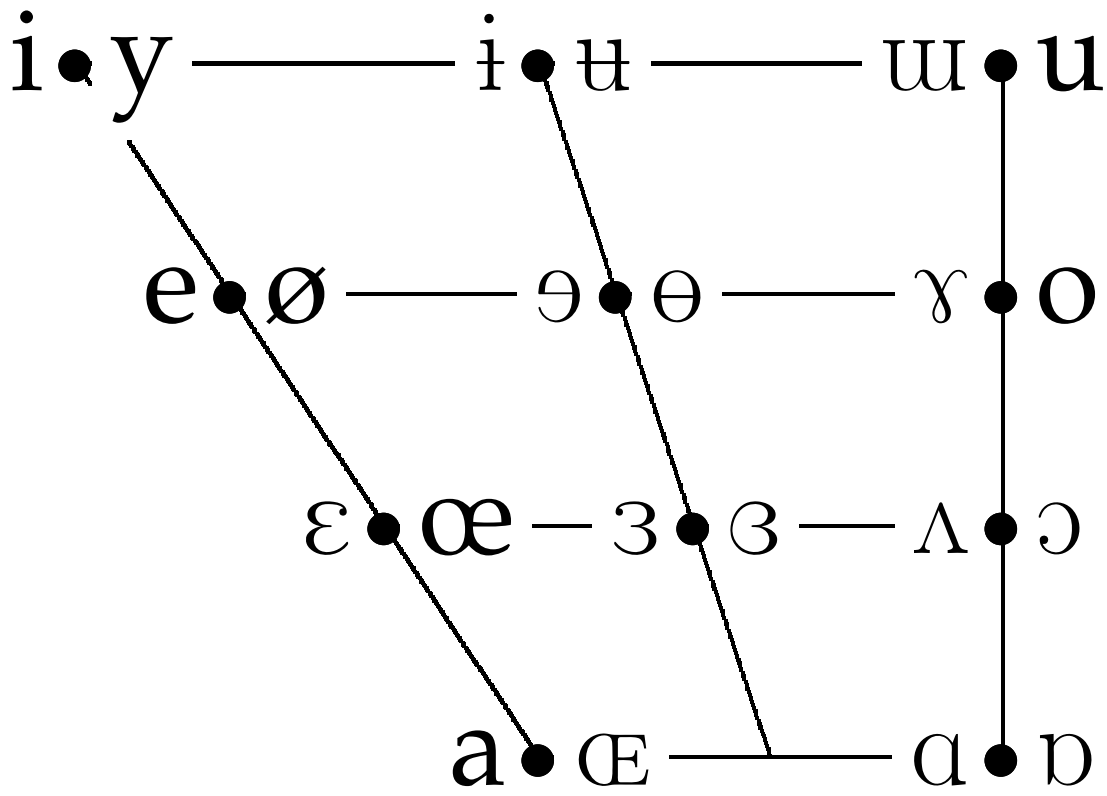
These articulatory characteristics are to be chosen according to the application.

Example: In the International Phonetic Alphabet (IPA) the cardinal vowels are represented with integers, the vowels in between with real numbers.

IPA Vowel diagram



IPA Vowel diagram



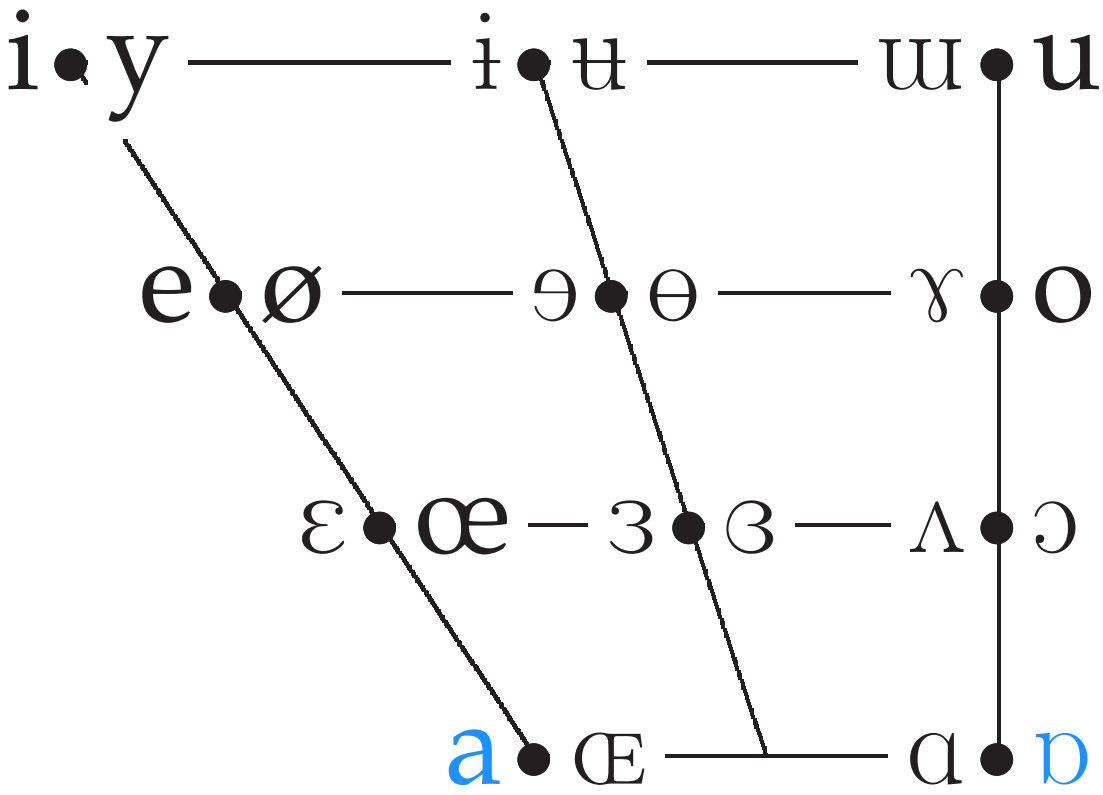
Signal Processing: Autocorrelation

Paula	*luki*	*lääkäri*	*-ksi*	*viide*	***-ssä***	*vuode*	***-ssa***.
Paula	las	Ärztin	als	5.	im	Jahr	im.

Paula schloss ihr Medizinstudium nach zehn Semestern erfolgreich ab.

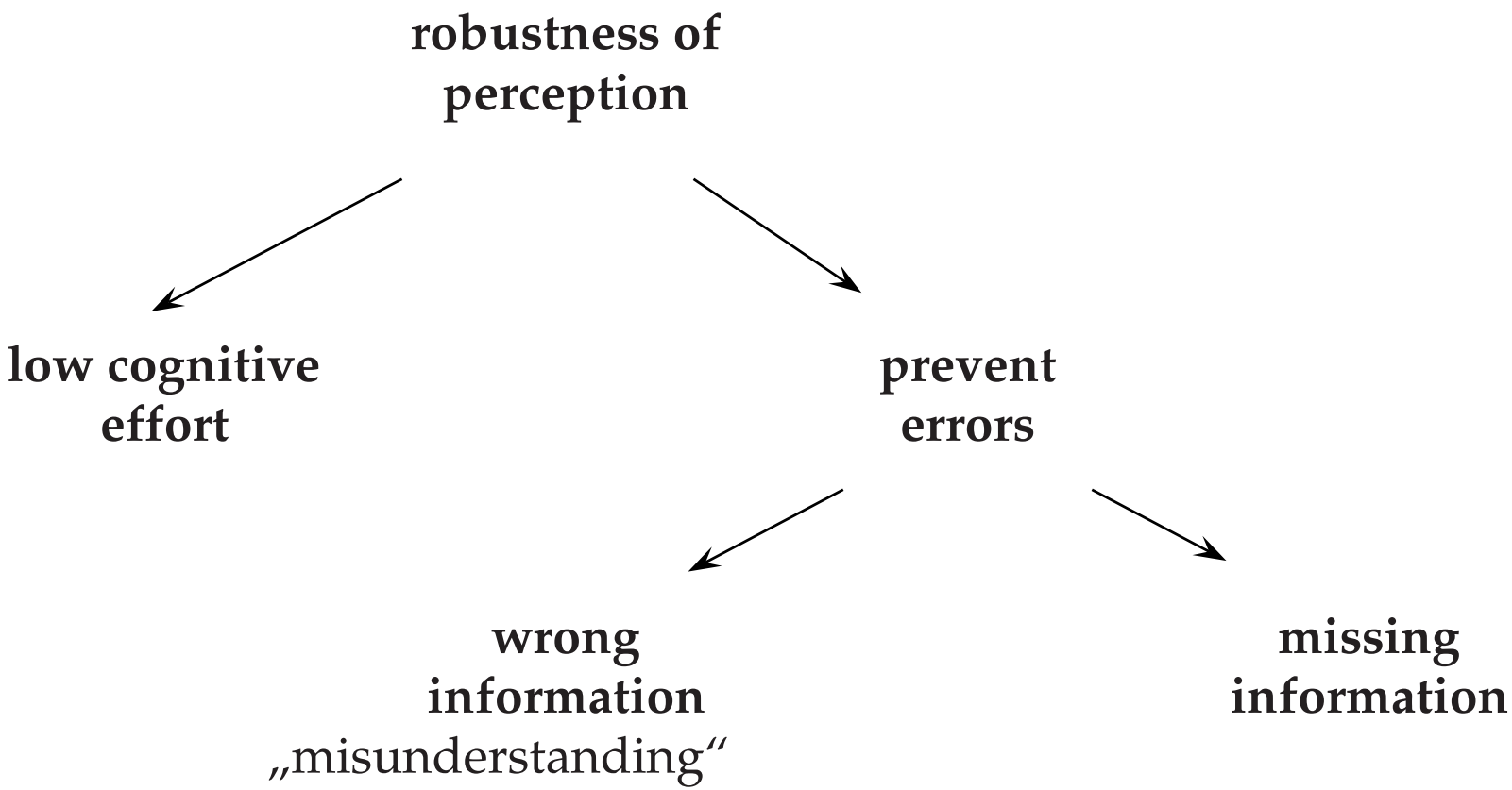
$\mathrm{a} = (3, 1, 1)$ und $\mathrm{ä} = (1, 1, 0)$

Signal Processing: Autocorrelation



Robustness Makes Natural Language Difficult to Learn!

Robustness


robustness of perception
low cognitive effort
prevent errors
wrong information
„misunderstanding“
missing information

Channel Codes

Automatic Repeat-reQuest (ARQ)

Detect errors and retransmit then.

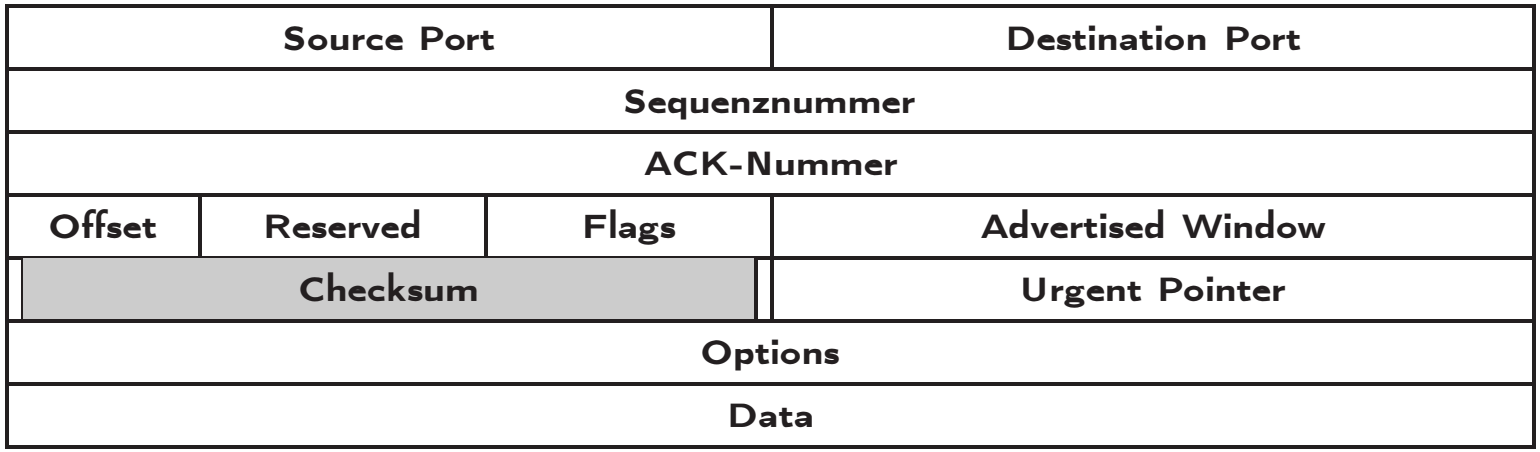
Forward-Error-Correction (FEC)

No retransmission necessary, possibility for error correction incorporated in first transmission.

Hybrid FEC/ARQ: Hybrid Automatic Repeat-reQuest (HARQ)

A hybrid method retransmits only if the error correction (FEC) didn't work but an error was detected (ARQ). This is called an **error-detection-and-correction** code.

Robustness in the TCP-Protocol



Robustness in Human Statement

Gestern habe ich in einem dick en Buch einen robust en Satz ge les en und als ich ihn so las , dacht e ich, ich könne mal . . .

- Place of the robustness not known without the data
- Amount of the robustness not known without the data

How Grammar is Formalized in Language Teaching

Generative Grammars

Ich springe auf **dem** Bett.

Ich springe auf **das** Bett.

No Analytical Grammars

There are (almost) no analytical grammars in language teaching!

Strange?!? A trivial insight for a theoretical computer scientist, but for centuries not for language teachers. They torture their pupils with the generative ones.

Theorem:

Every generative grammar can be converted to an analytical grammar and vice versa.

Declination of the German article

	m	**f**	**n**	**pl**
Nominativ	der	die	das	die
Genitiv	des	der	des	der
Dativ	dem	der	dem	der
Akkusativ	den	die	das	die

der	die	das	des	dem	den
5	4	2	2	2	1

Declination of the German article

f	Nominativ	die
f	Genitiv	der
f	Dativ	der
f	Akkusativ	die
m	Nominativ	der
m	Genitiv	des
m	Dativ	dem
m	Akkusativ	den
n	Nominativ	das
n	Genitiv	des
n	Dativ	dem
n	Akkusativ	das
p	Nominativ	die
p	Genitiv	der
p	Dativ	der
p	Akkusativ	die

Formal Grammar Presentation

Definition: *Formal Grammar Presentation*

A formal grammar $G = (V, \Sigma, R, S)$ formalizes a natural grammar. A formal grammar presentation formalizes the notation of a grammar in language teaching.

Formal Grammar Presentation

DEFINITION: *The Formal Grammatar Presentation G01*

The formal grammar ***presentation*** G01 of a grammar $G = (V, \Sigma, R, S)$, is a grammar $D = (V, \Sigma, R, S, X)$ augmented with a list of examples $X \in \Sigma^*$ and an order of the rules: R is not given as a set, but as a list.
V, Σ are S the same as in G.

Definition of the Grammar Process

Definition: *Grammar process*

A grammar process is the process of loosing or adding rules in a grammar G.
Let $R/R^- \cup R^+$, where R^+ is the set of rules which are added and R^- the set of rules which are lost.
Let $G_1 = (V, \Sigma, R, S)$ be the grammar before the process and G_2 the grammar which evolves from the process. Then

$$\text{Grammarprocess}(G_1) = (V, \Sigma, R/R^- \cup R^+, S) = G_2$$

If one or both of the sets R^+ and R^- are not empty then the process can be called a ***grammar change***.