

Wi froh bin i, daß i weg bin! Besta Frund,  
was isch das Herz des Menscha! ...

wIE froH bin ich, dAB ich weg pin! Oida! bEEster freund,  
Was ist Das herZ DeS mEnschen! ...

- Goethe: Die Leiden des jungen Werther

# Graphematische Transkription in Soziolekte

Vorarlbergerisch und kRocHa

## Ziele

- Praktische Präsentation einer CL-Anwendung
- Aufzeigen von Überlegungen, Problemen, Anregung zur Forschung
- Standardsprachliche Quelltexte in soziolektische Zieldtexte umwandeln
- Beispiele: Vorarlbergerisch und Krocha

## Übersetzungsprogramm?

- Ähnlich wie bei einem Übersetzungsprogramm
- Quell-Text wird zu Ziel-Text
- Bestimmte (sprachliche) Kriterien und Regeln müssen festgelegt werden
- Ähnliche Probleme und Ausgangsfragen

# Ausgangsfragen

- Was macht ein Soziolekt-Schreiber? Was muss eine Maschine können, um einen Soziolekt zu simulieren?
- Unterschiede der „Übersetzung“ (d.h. hier: Transkription) bei Mensch und Maschine
- Soziolinguistische Fragen (z.B. was wird durch das Schriftbild mitgeteilt?)
- Konzeption der Formalismen -> pragmatische Umsetzung
- Effizienter Aufbau des Systems
- Verfahren und Strategien um Texte zu transkribieren

5

# CL-Ansatz dieser Arbeit

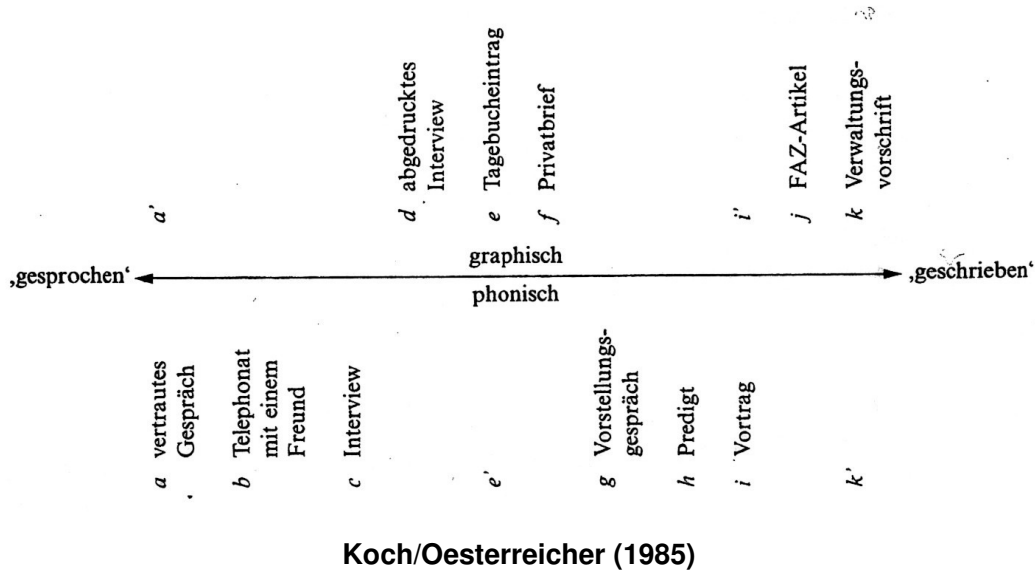
- technische Rekonstruktion (Präskription erforderlich)
- exakte Theorien vernachlässigt
- Funktionsfähigkeit im Vordergrund
- Pragmatik

6

# Theoretische und Pragmatische Überlegungen

# Besonderheiten der Chat-Kommunikation

# Hybrid zwischen Mündlichkeit und Schriftlichkeit



9

## Aspekte mündlicher Kommunikation/Nähe:

- Spontaneität
- Emotionalität
- Dialogizität
- Geringer Planungsaufwand
- Vorläufigkeit
- Kontextgebundenheit

10

## Aspekte schriftlicher Kommunikation/Distanz:

- Räumliche (nicht aber zeitliche!) Distanz
- Geringe Situations- und Handlungseinbindung
- Weitgehende Abwesenheit nichtsprachlicher Kontexte

## Weitere Einflüsse auf die Chat- Kommunikation:

- Zeitdruck
- Individueller Ausdruck und Gruppenzugehörigkeit

## Chat-spezifische Stilelemente:

- Orthographische und morphologisch-syntaktische Charakteristika
  - Phonetisch orientierte Schreibweise, Missachtung orthographischer Normen
  - Verkürzter sprachlicher Ausdruck
  - Reduzierte Interpunktion
  - Generelle Kleinschreibung

13

## Chat-spezifische Stilelemente:

- Pragmatische Charakteristika
  - Generelles Duzen zw. Unbekannten
  - Merkmale inszenierter Mündlichkeit
  - Graphostilistische Kompensationsstrategien nonverbaler Äußerungen
    - Iteration von Buchstaben
    - Majuskeln
    - Emoticons
    - Zustands- und Gefühlsäußerungen, Handlungsbeschreibungen
    - Onomatopoetika
  - Interjektionen

14

## Chat-spezifische Stilelemente:

- Lexikalische Charakteristika
  - Geringe Variation
  - Überwiegend kurze Lexeme
  - Anglizismen
  - Substandardregister und Dialekt

## Graphematik



# Graphematik

- Schriftlich umgesetzte Soziolekte
- Regeln treten in den Hintergrund
- Stark individuelle Schreibweisen
- Soziale Zugehörigkeit soll „erkennbar“ sein

17

# Graphematik

- Identifikation durch das Schriftbild
- Graphematische Umsetzungen technisch einfacher
- starke Vereinfachung durch Konzentration auf Graphematik
- aber: Gerechtfertigt durch pragmatische Orientierung
- folgende Umsetzungsregeln beruhen auf dem Schriftbild der Texte

18

# Effizienz

19

# Effizienz

- steht im Vordergrund der Anwendung
- Formulierung der Regeln möglichst effizient, d.h.
- Regeln sollen so allgemein wie möglich
- und speziell wie nötig sein
- Redundanz vermeiden
- Präskription vor Deskription

20

# Vorarlbergerisch

21

## Beispiel 1 – ‚sch‘

**XoOotRiStEsSeoOoX schreibt am 08.10.2008 08:05:**

hey du ! wenn kusch ez eigentlich genau noch wien ??  
... schriebsch ma denn a sms,wenn doa bisch ? ... hdl\*\*  
:)[...]

**unregistered schreibt am 05.11.2007 14:21 :**

kummscht wieder mol zu mir und denn können mir  
kuchen backen du woascht scho was i moan!!!! bis  
nächsten SAMSTAG!!!!!!!!!!!!!!

⇒ Regel: Immer wenn standardsprachlich –st kommt  
vorarlbergerisch –sch, selten auch –scht.

22

## Beispiel 2 – „en“ -> „a“

**Berger:** fotos aluaga mim luki!! [...]

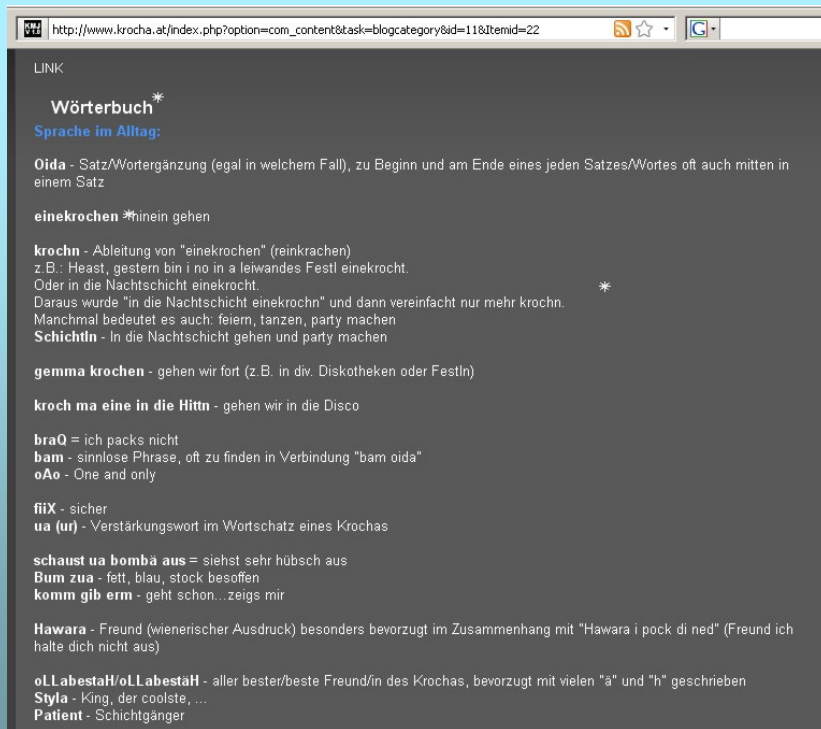
**kessi:** \*g\* der kuacha luagt geil us. merci [...]

**kessi:** jo... bis winächta isch as eba ziemlich viel [...]

**matze:** he kessi tuan ihr do duna eigatli o studiera oder nur sufa ;-) ???

⇒ Regel: Immer wenn standardsprachlich Endsilbe ‚-en‘ kommt vorarlbergerisch ‚a‘

Krocha



25

## Probleme beim Krocha-Korpus

- Uneinheitliches, heterogenes Schriftsystem
- kein verschriftlichter mündlicher Dialekt
- beruht tatsächlich v.a. auf dem Schriftbild
- d.h. je nach Herkunft andere Schreibweise
- zudem: stark intentionell geleitet (Ankündigung, vs. allgemeine Diskussion)
- daraus ergeben sich verschiedene Grade der Text-Inszenierung

26

# Regeln von krocha.at

- 1. Oida - Satz/Wortergänzung (egal in welchem Fall), zu Beginn und am Ende eines jeden Satzes/Wortes oft auch mitten in einem Satz
- 2. bam - sinnlose Phrase, oft zu finden in Verbindung "bam oida"
- 3. abwechselnde Groß- und Kleinbuchstaben
- 4. z -> c
- 5. i -> y
- 6. a -> ä
- 7. u -> ü
- 8. b -> p
- wichtige Vokabeln:
  1. hinein gehen -> einekrochen
  2. gehen wir fort/weg -> gemma krochen
  3. sicher -> fix
  4. ur -> ua

27

## Anpassung der Regeln nach Korpusvergleich

- wichtig va. „bam“ und „oida“
- allerdings fast ausschließlich an Satzanfang und –ende
- Buchstabenersetzungen eher selten; außer bei starkem Inszenierungsgrad
- Groß-/Kleinschreibung nicht „abwechselnd“

28

# Informatische Umsetzung

29

## Verschiedene Ansätze

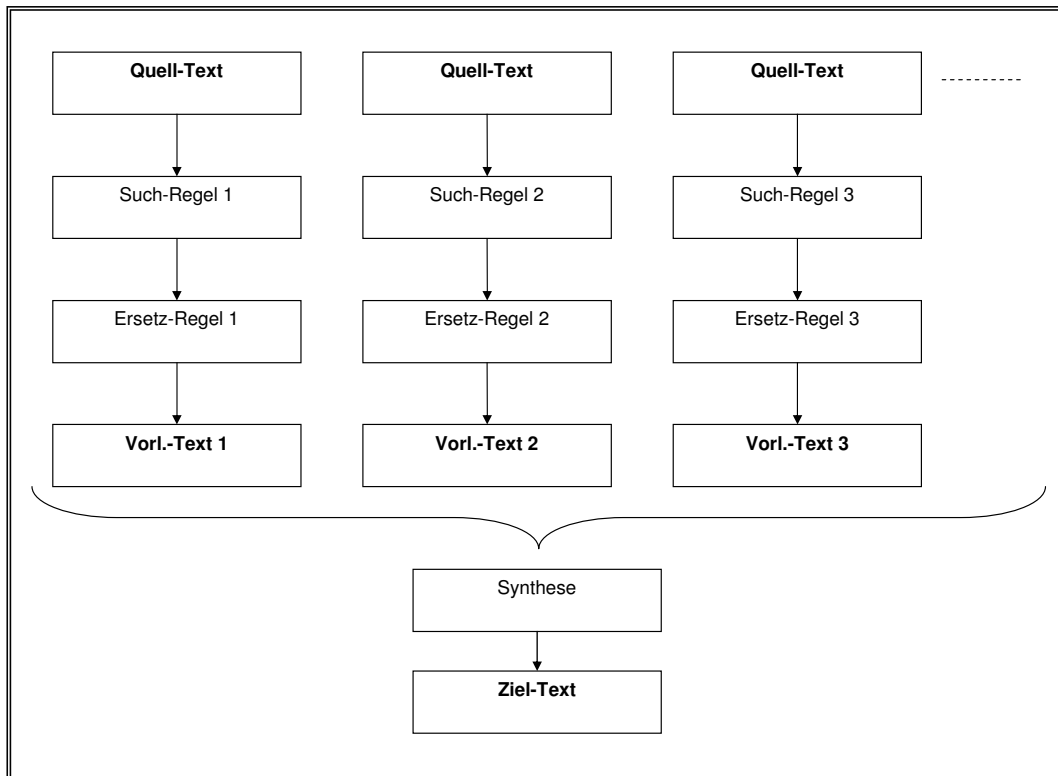
- Grammatikalisch vs. Graphematisch
- Übersetzung vs. Transkription
- Parsing vs. Suchen-und-Ersetzen

## Graphematische Modelle

- Paralleles Modell
- Lineares Modell

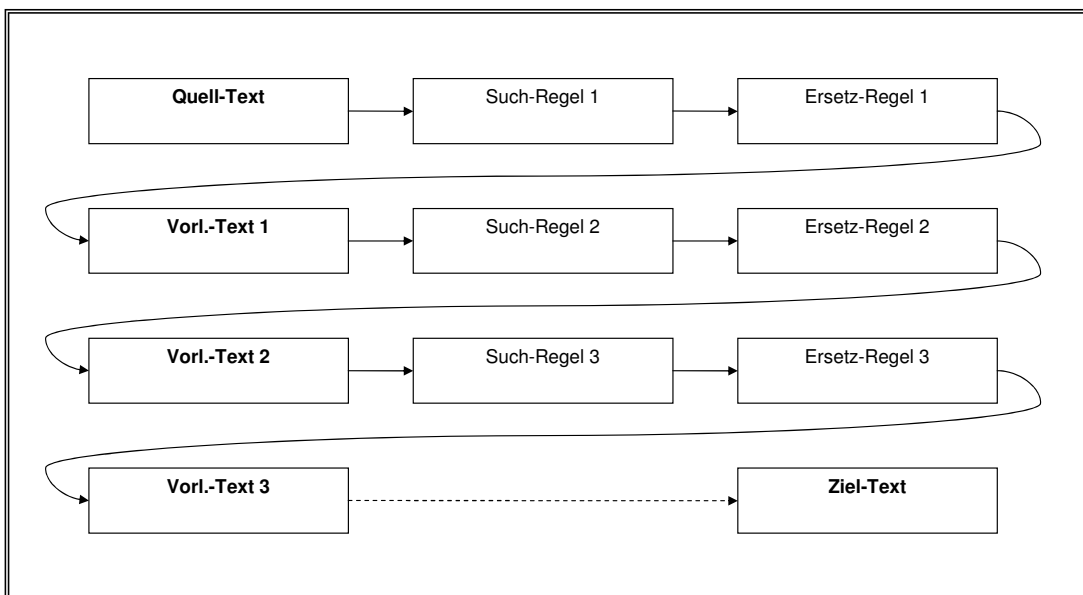
30

## Paralleles Modell



31

## Lineares Modell



32



# Umsetzung

- Suchen und Ersetzen
- Lineares Modell
- basierend auf PHP
- Anpassungsfähige Oberfläche
- Zufallsgenerator
- Auswertungsfunktion