

SCHRIFTSYSTEME, TRANSLITERATION UND TRANSKRIPTION

SEMINAR: SUCHMASCHINEN UND RAG

DOZENT: STEFAN LANGER

CIS, UNIVERSITÄT MÜNCHEN

SOMMERSEMESTER 2024

Einleitungsfragen

Welche Schriftsysteme kennen Sie? Finden Sie jeweils ein Beispiel für jeden der unten angegebenen Typen

- Alphabete
- Konsonantenschriften (engl *consonantary* oder *abjad*)
- Abugida (Zwischending zwischen Silbenschrift und Alphabet)
- Silbenschriften
- logographische Schriftsysteme

Typen von Schriftsystemen

Alphabete

- Lateinische Schrift
- Griechische Schrift
- Kyrillische Schrift

Konsonantenschriften (engl *consonantary* oder *abjad*)

- Arabische Schrift
- Hebräisch

Abugida (Zwischending zwischen reiner Silbenschrift und Alphabet)

- Devanagari (für Hindi u.a.) + die meisten indischen Schriftsysteme)

Silbenschriften

- Kana (Katakana/Hiragana; Japanisch)

logographische Schriftsysteme

- Chinesisch (und Ableger, z.B. japanische Kanji)

Alphabet

In einem Alphabet werden alle Phoneme einer Sprache durch Zeichen ausgedrückt (also Vokale und Konsonanten). Ein Alphabet ist damit eine phonemische Wiedergabe einer Sprache.

Einschränkungen:

Lautveränderungen, die nicht in der Schrift wiedergegeben sind (Englisch!)

Bsp. "light" vgl. dt. "Licht"

Laute können auch durch Kombinationen von Zeichen wiedergegeben werden (z.b. Deutsch *sch*, English *sh*)

Verhältnis Laut-Zeichen kann auch sein 0:N N:0

N:0: z.B. Frz. *Écout***ent** ([ekut])

0:N: z.B. Deutscher Knacklaut/Glottalverschluss (ver'eisen)

Schreibung in Alphabetschriften

In Alphabetschriften werden unterschiedliche Wortformen in der Regel auch in der Schreibung ausgedrückt

Teilweise werden in der Schreibung morphologische Regularitäten erhalten, die es in der gesprochenen Sprache nicht mehr vorhanden sind:

Frz.: donner, **donne**, **donnes**, **donne**, donnons, donnez, **donnent**

Konsonantenschrift

In einer Konsonantenschrift werden alle Konsonanten einer Sprache durch Zeichen ausgedrückt (Vokale nicht).

Einschränkungen:

Das Arabische z.B. schreibt keine Kurzvokale, aber Langvokale (allerdings durch Buchstaben, die auch für Konsonanten stehen können)

القمر

Al-q(a)m(a)r

Diakritika können optional verwendet werden, um die kurzen Vokale anzuzeigen

القَمَرُ

Schreibung in Konsonantenschriften

In Konsonantenschriften werden unterschiedliche Wortformen in der Regel auch in der Schreibung ausgedrückt, wenn sie Konsonanten betreffen; Vokaländerungen sind in der Schrift nicht sichtbar

Abugida (Untertyp einer Silbenschrift)

Silbenschrift mit Elementen der Alphabetschriften

Beispiele Devanagari (indische Schrift v.a. verwendet für Hindi), indische Schriften, Thai
(Abbildung: Devanagari aus wikipedia.de)

Vokale				
Grundstufe		Dehnstufe		
अ a	क ka	आ ā	का kā	
इ i	कि ki	ई ī	की kī	
उ u	कु ku	ऊ ū	कू kū	

Vokale werden dargestellt, indem die Konsontenzeichen modifiziert werden. Im Gegensatz zu einer reinen Silbenschrift lässt sich also im Zeichen, das die Silbe repräsentiert, der Konsonantenanteil noch isolieren.

Morphologie / Schreibung in Abugidas

Hier gilt ähnliches wie für Alphabetschriften

Silbenschriften

Jedes Zeichen
steht für eine
ganze Silbe

V.a. Japanisch
(Hiragana,
Katakana) und
einige ältere bzw.
ausgestorbene
Sprachen.

Beispiel: Hiragana

	<i>a</i>	<i>i</i>	<i>u</i>	<i>e</i>	<i>o</i>
Ø	<u>あ</u>	<u>い</u>	<u>う</u>	<u>え</u>	<u>お</u>
<i>K</i>	<u>か</u>	<u>き</u>	<u>く</u>	<u>け</u>	<u>こ</u>
<i>S</i>	<u>さ</u>	<u>し</u>	<u>す</u>	<u>せ</u>	<u>そ</u>
<i>T</i>	<u>た</u>	<u>ち</u>	<u>つ</u>	<u>て</u>	<u>と</u>
<i>N</i>	<u>な</u>	<u>に</u>	<u>ぬ</u>	<u>ね</u>	<u>の</u>
<i>H</i>	<u>は</u>	<u>ひ</u>	<u>ふ</u>	<u>へ</u>	<u>ほ</u>
<i>M</i>	<u>ま</u>	<u>み</u>	<u>む</u>	<u>め</u>	<u>も</u>
<i>Y</i>	<u>や</u>		<u>ゆ</u>		<u>よ</u>
<i>R</i>	<u>ら</u>	<u>り</u>	<u>る</u>	<u>れ</u>	<u>ろ</u>
<i>W</i>	<u>わ</u>	<u>ゐ</u>		<u>ゑ</u>	

Logographische Schriftsysteme

Bei logographische Schriftsystemen stehen die Zeichen nicht für einzelne Laute, sondern für ganze Morpheme oder Wörter

Logographisch Schriftsysteme haben oft ideographische und piktographische Aspekte; d.h. die Zeichen stehen direkt für Begriffe oder haben bildhaften Bezug zum dargestellten Begriff

森林

Logographische Schriftsysteme

Beispiele:

Das Chinesische ist ein (weitgehend) logographisches Schriftsystem. Da das Chinesische eine isolierende Sprache ist, besteht keine Notwendigkeit Flexionssuffixe in das System zu integrieren

Das Japanische verwendet Chinesische Schriftzeichen, häufig aber in anderer Bedeutung als im Chinesischen. Flexionsendungen werden durch Hiragana-Zeichen (Silbenschrift) dargestellt

Groß- und Kleinschreibung

Manche Schriftsysteme unterscheiden zwischen Groß- und Kleinschreibung

- Lateinische Schriftzeichen
- Griechisch

Großschreibung für bestimmte Wortarten; Satzanfang; Markierung (Majuskeln)

Generelle Problem der Groß- Kleinkonversion

- Erhöhung der Ambiguität

Spezielle Probleme der Groß- Kleinkonversion

- Buchstabenersetzungen
 - GRÜSSE $\leftrightarrow$ Grüße
- Akzentwegfall, z. B. Französisch
 - *égalité* – *Egalité* (selten *Égalité*)
- Sprachenspezifische Groß-Kleinkonversion für einzelne Buchstaben
 - Türkisch I $\leftrightarrow$ ı İ $\leftrightarrow$ i (normalerweise I $\leftrightarrow$ i)

Diakritika

Diakritika sind Zeichen, die zu Buchstaben hinzutreten; z.B. Akzentzeichen im Französischen.

- Die Abgrenzung von Buchstaben mit Diakritika zu eigenständigen Buchstaben ist schwierig

Diakritika werden in vielen Sprachen inkonsistent verwendet

Eventuell werden Diakritika nur auf Kleinbuchstaben gesetzt.

Diakritika werden bei Wiedergabe in anderen Sprachen gerne weggelassen

- ähnliches gilt für sprachspezifische Buchstaben wie Ø
- tastaturbedingt

Diakritika können unterschiedlich realisiert werden (zusammengesetzte Buchstaben vs. getrennte Realisierung (combining diacritical marks))

- → Normalisierung erforderlich

Ligaturen

Ligaturen sind zusammengesetzte Buchstaben, z.B. *ff*, *fi*, *fl*, *ft*

Es gibt auch historische Ligaturen, die zu eigenständigen Buchstaben geworden sind, z.B. w, ß

Bei der Normalisierung werden Ligaturen in der Regel aufgelöst und die Einzelbuchstaben

Schriftrichtung

Links nach rechts (europäische Sprachen)

Rechts nach links (arabische Schrift, Hebräisch)

...

القمر

Al-q(a)m(a)r

TRANSLITERATION UND TRANSKRIPTION

MASTERSEMINAR SUCHMASCHINEN

STEFAN LANGER, CIS, UNIVERSITÄT MÜNCHEN

SOMMERSEMESTER 2023

Transkription

Was ist Transkription?

وليم شكسبير (ar)

ويليام شكسبير (ar)

Ουίλλιαμ Σαίξπηρ (el)

Уильям Шекспир (ru)

Вільям Шекспір (uk)

วิลเลียม เชกสเปียร์

Transliteration und Transkription

Transliteration ist die Übertragung eines Texts/Worts von einem Schriftsystem (meist dem normalerweise für eine Sprache benutzten Schriftsystem) in ein anderes; beispielsweise die Darstellung kyrillischer Texte in lateinischer Schrift;

- Bei einer wissenschaftlich exakten Transliteration wird dabei weitgehend buchstabengetreu und reversibel übertragen (allerdings ist die Übertragung aus Konsonantenschriften wie dem Arabischen meist nicht 100% reversibel)

Transkription ist eine meist lautbasierte Übertragung in ein anderes Schriftsystem

Retransliteration ist die Rückübertragung in das ursprüngliche Schriftsystem

Transliteration/Transkription wird am häufigsten für Eigennamen angewandt (da diese in der Regel nicht übersetzt werden können).

Sprache?

EshTaaGt LeeL 7eeLM E93'eeR .. WJDaaR Min 6eeN O 7a9eeR

Eshtagt lel 7eelm el93'eer ..

Wjdar men 6en o 7a9eer ..

WGamra wara el leeel el6'reer ..

Varianten

Anwar El Sadat (Arabic: محمد أنور السادات **Anwar as-Sādāt**, Egyptian Arabic pronunciation: [mæ'hæmmæd 'ʔanwarʕ essæ'dæ:t]; 25 December 1918 – 6 October 1981) was the third President of Egypt, serving from 15 October 1970 until his assassination by fundamentalist army officers on 6 October 1981. In his eleven years as president, he changed Egypt's direction, departing from some of the economic and political principles of Nasserism by re-instituting the multi-party system, and launching the Infitah economic policy.

Aus hr.wikipedia:

Anvar el-Sadat

Muhammad Anwar Al Sadat (Abul-kum, 25. prosinca 1918. - Kairo, 6. listopada 1981.), egipatski vojnik i predsjednik.

Transliteration und Transkription in Suchmaschinen

Mehrsprachige Datenvorhaltung

Erkennung und Normalisierung transkribierter Entitäten
(Personennamen, Ortsnamen)

Blogs, SMS u.a. Texte in Transkription erkennen und durchsuchen