Sprachsynthese: Graphem-Phonem-Konvertierung

Uwe Reichel
Institut für Phonetik und Sprachverarbeitung
Ludwig-Maximilians-Universität München
reichelu@phonetik.uni-muenchen.de

8. Januar 2021

Inhalt

- Einflussfaktoren
- Alinierung
- Konvertierung
  - Table Lookup with Defaults (van den Bosch et al., 1993)
  - Maschinelles Lernen
  - Entscheidungsbäume
- Silbifizierung
  - in Graphemfolge
  - in Phonemfolge
- Wortbetonung

Einflussfaktoren

Notation: *Grapheme: in spitzen Klammern; Phoneme' in Schrägstriche nach SAM-PA; Graphem/Phonem/Wortvariablen: kursive Kleinbuchstaben; Graphemklassen (Vokal): kursive Großbuchstaben (V);*

$\_\_\#_\sigma] =$ *'vor Silbengrenze'*

- Ist die Abbildung $<Graphem> \longrightarrow /Phonem/$ eindeutig?
- **Nein**. Beispiel $<s>$:

  $<$s$>$ $\longrightarrow$ /s/ *(was)*
  $<$s$>$ $\longrightarrow$ /z/ *(Vase)*
  $<$s$>$ $\longrightarrow$ /S/ *(stehen)*
  $<$s$>$ $\longrightarrow$ /_/ *(Wasser)*

- weiteres Beispiel: $<u>$ in $<B\mathbf{u}nd>$ und $<Q\mathbf{u}alle>$

Einflussfaktoren

Graphemkontext

<s> $\longrightarrow$ /z/ | *V _V* ; **aber:** *losen* **vs.** *Loserwerb*
<s> $\longrightarrow$ /S/ | _<t>; **aber:** *Stabilität vs. Rost*

. . . reicht nicht aus

Silbenstruktur

- Auslautverhärtung, keine Beeinflussung durch Graphemumgebung über Silbengrenzen hinweg
- <*g*> $\longrightarrow$ /*g* | *k*/: *Wege*, *Weg*
- <*s*> $\longrightarrow$ /*z* | *s* | *S*/: *Vase*, *Häuschen*

Einflussfaktoren

Morphologie

- morphologischer Einfluß direkt und über die Silbenstruktur manifestiert
- **direkt:** Phonem-Identität abhängig von Morphemklasse
- **Beispiele:**
  - <*er*> in *Erlöser* (/?E6l2:z6/); morph. $er_{prefix} + lös_{verb} + er_{suffix}$.
    <*er*> $\longrightarrow$ /?E6/ | im Präfix
    <*er*> $\longrightarrow$ /6/ | im Suffix
  - <*e*> in *geben* (/ge:b@n/); morph. $geb_{verb} + en_{infl}$.
    <*e*> $\longrightarrow$ /e:/ | im einsilbigen Verbstamm
    <*e*> $\longrightarrow$ /@/ | in Flektionsendung

Einflussfaktoren

- **indirekt:** morphologische Struktur bestimmt Silbenstruktur und damit Phonem-Identität
- **Beispiele:**
  - <*ng*> in *Angel* (/?aN@l/) vs. *Angelegenheit* (/?ang@le:g@nhaIt/):
    morph. $angel_{noun}$ vs. $an_{prefix}+ge_{prefix}+leg_{verb}+en_{suffix}+heit_{suffix}$.
    <*ng*> $\longrightarrow$ /N/ | keine Silbengrenze
    <*ng*> $\longrightarrow$ /n.g/ | (hier: morphologisch bedingte) Silbengrenze
  - <*se*> in *losen* (/lo:z@n/) vs. *Losentscheid* (/lo:s?EntSaIt/):
    morph. $los_{verb}+en_{infl}$ vs. $los_{noun}+ent_{prefix}+scheid_{verb}$.
    <*se*> $\longrightarrow$ /z@/ | keine Silbengrenze
    <*se*> $\longrightarrow$ /s.?E/ | (hier: morphologisch bedingte) Silbengrenze

Alinierung: Grundlagen

- Zur Gewinnung des G2P-Trainingsmaterials
- Zuordnung zusammengehöriger Abschnitte in Graphem- und Phonemsequenzen

O	b	e	r	s	c	h	u	l	z	e	u	g	n	i	s
?+o:	b	6	__	S	__	__	u:	l	t+s	OY	__	k	n	I	s

- Alinierungsproblem formuliert als **Minimierung der Distanz zwischen Graphem- und Phonemsequenz**

Alinierung: Levenshtein-Distanz

Levenshtein-Distanz

- **Minimal nötige Editierkosten** um Sequenz v (Grapheme) in Sequenz w (Phoneme) umzuwandeln
- **Standard-Editieroperationen:**
  - **Substitution** von v_i durch w_j: $<u> \longrightarrow /u:/$
  - **Löschung** von v_i: $<r> \longrightarrow \_$
  - **Einfügung** von w_j: $\_ \longrightarrow /?/$

Alinierung: Kostenfunktion

Naiv

- **Substitution**

$$c(v_i, w_j) = \begin{cases} 0 & : \quad v_i == w_j \\ 1 & : \quad \text{else.} \end{cases} \tag{1}$$

- **Löschung, Einfügung**

$$c(v_i, \_) = 1 \tag{2}$$

$$c(\_, w_j) = 1 \tag{3}$$

- brauchbar für Wortvergleiche (z.B. automatische Rechtschreibkorrektur)
- unbrauchbar für Graphem-Phonem-Alignment: <x> $\neq$ /x/

Konvertierung: Table Lookup with Defaults

Table Lookup with Defaults (van den Bosch et al., 1993)

- **Training:** Speicherung des jeweils kürzesten Graphemkontexts für ein eindeutiges Graphem-Phonem-Mapping in Tabelle G

k **a** m m	/a/
u ß	/u:/

- **Sortierung** nach Länge der Graphemsequenz
- **2 Default-Tabellen:** Graphem-Fenster + am häufigsten damit ko-okkurierendes Phonem

$<v_{i-1}><v_i><v_{i+1}>$	$/w_j/$
$<v_i>$	$/w_j/$

Konvertierung: Table Lookup with Defaults

- **Konvertierung:**
  - Suche nach passendem Graphem-Muster (von lang nach kurz) in Tabelle G
  - Falls nicht vorhanden, Rückgriff auf Default-Tabellen
  - Beispiel:
    - zu konvertieren: $<u>$ in *Fuß*
    - in Tabelle G gefundenes Muster: *uß*

    $\longrightarrow$ Ausgabe: */u:/*
- **Vorzüge**
  - rein datenbasierter Ansatz

  $\longrightarrow$ kein Expertenwissen nötig, sprachunabhängig

Maschinelles Lernen

- **Ziel:** Erlernen des Zusammenhangs zwischen Zielwerten (Kategorien oder kontinuierliche Werte) für Objekte und deren Attributen.
- bezogen auf Graphem-Phonem-Konvertierung
  - **Objekte:** Grapheme
  - **Attribute:** Graphem-Identität, umgebende Grapheme, Position des Graphems innerhalb der Silbe, ...
  - **Zielwerte:** Phoneme

Maschinelles Lernen

- Objekte als **Attributbündel** (*Feature vectors*) repräsentiert. Beispiel:
  - **Attribute** für Graphem v_i:
    $[<v_{i-1}>,<v_i>,<v_{i+1}>$, Morphemtyp, $\_\#_\sigma]$
  - Mögliche Attributwerte: [a–z, a–z, a–z, *frei*|*gebunden*, 0|1]
    (*gebundenes* Morphem kann allein kein Wort bilden)
  - Feature vector für erstes $<e>$ in *g***e***ben*:
    [*g, e, b, gebunden, 1*]
  - **Zielwert (hier kategorial):** */e:/*

Maschinelles Lernen

- **Attribute:** kategorial oder kontinuierlich
  - **kategorial:** Graphem-Identität, Position in Silbe, Phonemklasse, Wortbetonung, Morphstatus
  - **kontinuierlich:** relative Position des Graphems im Wort, Lautdauer, F0-Wert

Maschinelles Lernen

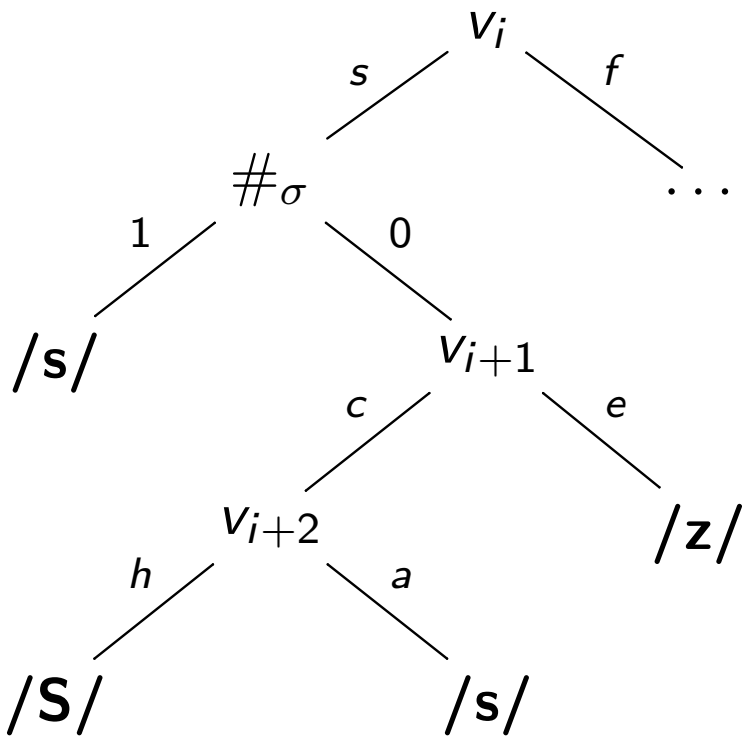
- **Überwachtes Lernen:** Zielwerte in Trainingsdaten bekannt; C4.5, CART, neuronale Netze (ANN)
- **Unüberwachtes Lernen:** Zielwerte nicht bekannt; Clustering, ANN
- **Versch. Methoden für Attribut- $\longrightarrow$ Zielwert-Typen:**
  - C4.5: kategorial/kontinuierlich $\longrightarrow$ kategorial; z.B. Akzent
  - CART: kategorial/kontinuierlich $\longrightarrow$ kategorial/kontinuierlich; z.B. Lautdauer
  - ANN: kontinuierlich $\longrightarrow$ kategorial/kontinuierlich

Entscheidungsbäume

- Quinlan (1993); http://www.cse.unsw.edu.au/~quinlan
- **Modellierung:**
  - **Objekte:** Pfade durch den Baum
  - **Attribute:** Kanten zwischen Knoten
  - **Zielwerte:** terminale Knoten
- Vorteil: Transparenz $\longrightarrow$ Wissensakquirierung möglich

Entscheidungsbäume

Beispielausschnitt:



Entscheidungsbäume

Rekursiver Aufbau des Baums

- **Fall 1:** Ordne alle Objekte, die noch nicht durch einen vollständigen Pfad im Baum repräsentiert sind und den **gleichen Zielwert** haben, einem neuen Blatt (terminaler Knoten) zu.
- **Fall 2:** Genauso, wenn die Objekte zwar **verschiedene Zielwerte** haben, sich **aber** anhand der gegebenen Attribute **nicht mehr weiter unterscheiden** lassen.
- **Fall 3:** Haben Objekte **verschiedene Zielwerte und unterscheiden** sie sich in einer oder mehreren Attributen, so wähle das **zur Partitionierung der Objektmenge am 'besten geeignete' Attribut** und erzeuge einen Knoten, an dem sich der Baum in mehrere durch Werte des betrachteten Attributs vorgegebene Kanten aufspaltet.

Entscheidungsbäume

G2P-Anwendung von Entscheidungsbäumen

- **Attribute:**
  - Graphemkontext (n Grapheme vorher/nachher)
  - Morphologie: Morphemklasse, +/– folgende Morphemgrenze
  - Silben: Aufbau der Silbe Onset/kein Onset (bedeckt/nackt), Coda/kein Coda (geschlossen/offen), Position in Silbe (Onset, Nukleus, Coda, Gelenk *alle*)
  - Phonem: Vorgeschichte (n Phoneme vorher)
- **Zielwert**: Phonem, incl. leeres Phonem (/_/), oder Phonem-Cluster (/?+a:/)

Morphologische Zerlegung

Voraussetzung: Morphem-Lexikon mit Morphemklassifizierung

1. Teile jedes Wort *w rekursiv* von links nach rechts in String-Präfixe und -Suffixe bis eine erlaubte Segmentierung möglich ist oder das Wortende erreicht wird.
2. Eine Grenze, die den aktuellen String in Präfix und Suffix unterteilt dann akzeptiert wenn (i) das Präfix im Lexikon zu finden ist, und (ii) eine erlaubte Segmentierung des Suffixes möglich ist, und (iii) die Kombination 'Präfix-Klasse + Klasse des ersten Suffix-Teils' nicht der Morphotaktik widerspricht, und (iv) die Klasse des letzten Suffix-Teils kompatibel ist mit dem POS von *w*.

Morphologische Zerlegung

Beispiel: *Fassade – nkletterer*

(i) String-Präfix *Fassade* ist im Lexikon

(ii) **Rekursion:** erlaubte Segmentierung des String-Suffixes *nkletterer* möglich (*n-kletter-er*)

(iii) Morphemklassen *Fassade/NN – n/Fugenmorphem* kompatibel

(iv) Morphemklasse des letzten String-Suffixes *er/NN-Suffix* kompatibel mit POS NN des Worts

$\longrightarrow$ Segmentierung *Fassade – nkletterer* möglich

Silbifizierung

In Graphemfolge

- **3 vorherzusagende Klassen:**
  Silbengrenze folgt, folgt nicht, Ambisyllabizität
- **Feature-Auswahl:**
  - für jedes Graphem $<v_i>$
  - innerhalb eines auf $<v_i>$ zentrierten symmetrischen Graphem-Fensters
  - Graphem, Konsonant/ Vokal
  - ggf. Morphemgrenze (+/− relevant für Silbengrenze)

Silbifizierung

- für Silbifizierung **relevante morphologische Grenzen**:
  - vor allen Morphemen außer Flexionsendungen, Suffixen, Komparationsmorphemen und Fugen
  - vor Flexionsendungen, Suffixen mit initialem Konsonanten und eigenem Silbenkern (*schaff***te**)
  - vor Flexionsendungen, Suffixen mit initialem Vokal, wenn das vorangehende Morphem auf Vokal endet (*bau***en**)

Erweiterungen

Folgende Folien sind *kein Prüfungsstoff* bis Abschnitt "Evaluierung".
Nach abgeschlossener G2P Konvertierung Erweiterung durch:

- Phonetische Sibifizierung
- Wortbetonung (*lexical accent*)

Silbifizierung

In Phonemfolge

1. setze vor jedes Sonoritätsminimum eine Silbengrenze
2. Feinadjustierung gemäß Kohlers (1995) Silbenphonotaktik und silbengrenzrelevanter Morphemgrenzen

Beispiel:

/fE6hEltnIs/ $\xrightarrow{1.}$ /fE6.hEl.tnIs/ $\xrightarrow{2.}$ /fE6.hElt.nIs/

Phonetische Silbifizierung

Silbenphonotaktik (Kohler, 1995)

$$\left(\left\{ \begin{array}{ccc} & K_{a,b,c} & \\ (K_a) & K_a & K_b \\ & K_a & K_c \\ (K_a) & K_a & K_a \end{array} \right\} \; V \; \left\{ \begin{array}{ccc} & K_{a,b} & \\ K_b & K_a & (K_a) \\ K_b & K_b & (K_a) \\ K_a & K_a & \end{array} \right\} \right) \left(\left\{ \begin{array}{c} K_a(+K_a) \\ +K_a(K_a) \end{array} \right\} \right)$$

- K_a: Plosive, Frikative
- K_b: Nasale, /l/, /r/
- K_c: /h/, /j/
- V: Vokale
- $+$: Morphemgrenze

Phonetische Silbifizierung

Silbenbeispiele:

- *Herbsts* /hE6psts/ $K_c V K_a K_a + K_a$
- *Psalm* /psalm/ $K_a K_a V K_b K_b$

Restriktionen gegen Übergeneralisierung

- **Beispiel:** für $(K_a)K_a K_b$ darf K_a nicht im Artikulationsort mit K_b übereinstimmen,
  vgl. /fE6.hEl.tnls/ $\longrightarrow$ /fE6.hElt.nls/

Wortbetonung

Simplex-Wörter: Restriktionen

- **Drei-Silben-Fenster**: Wortbetonung kann nur auf eine der letzten drei Silben im Wort fallen (Ultima, Penultima, Antepenultima)
  Ausnahme: *schwa* in Antepenultima (/ 'a:.b@n.tOY.6 /)
- **Closed penult**: geschlossene Penultima verhindern eine Betonung weiter links (/ hi.b'Is.kUs /)
- **Final schwa**: Betonung der Penultima, wenn der Silbenkern der Ultima aus einem Schwa besteht: / tsi.tr'o:.n@ /;
  Produktivität: / g'e:.nE.zIs / $\longrightarrow$ / gE.n'e:.z@ /

Wortbetonung

Allgemeine Tendenzen im Standard-Deutschen

- Wortbetonung eher hinten im Wort
- eher auf *schweren* Silben (Langvokal/ Diphtong, Coda)
- *schwa*-Silben nicht betonbar

Wortbetonung

Simplex+Affixe

- **betonte Morphemklassen:**
  - Verbpartikeln: **w'eg***fahren*
  - betonte Affixe: *Abstin'***enz**, *pass'***abel**
- **unbetonte Morphemklassen:**
  - Flektionsendungen
  - unbetonte Affixe: **ent***spr'echen*, *M'ann***schaft**, *'Ärge***mis**

Komposita

- häufig Haupt- und Nebenbetonung
- **zweigliedrig:** Hauptbetonung auf erstem Glied (wenige Ausnahmen: *Lebew'ohl*)

Wortbetonung

- **mehrgliedrig:** Anwendung der **Compound-stress-Rule (CSR)** der metrischen Phonologie:

  CSR: *im Kompositum AB ist B strong s, wenn es sich weiter verzweigt, ansonsten ist A strong und B weak w*

- Erklärung von **Ausnahmen:** Atomisierung lexikalisierter verzweigender Konstitutenten: *K'unst#[hand#werk]*, daher: drei– $\longrightarrow$ zweigliedrig $\longrightarrow$ Hauptbetonung auf erstem Glied

Wortbetonung

weitere Schwierigkeiten

- **stress shift**: *D'oktor* > *Dokt'oren* (vgl. *Final schwa*-Restriktion); *'Ablauf* > *Progr'amm#abl"auf*
- **Homographen** unterschiedlicher Wortart, Valenz
  - *mod'ern/ADJ* **vs.** *m'odern/V*
  - *K'onstanz/NE* **vs.** *Konst'anz/NN*
  - *'allerhand/Indefpron,ADV* **vs.** *allerh'and/ADJD*
  - *d'arüber/PAV* **vs.** *dar'über/PTKVZ*
  - *d'urchlaufen/V(intrans)* **vs.** *durchl'aufen/V(trans)*

Wortbetonung

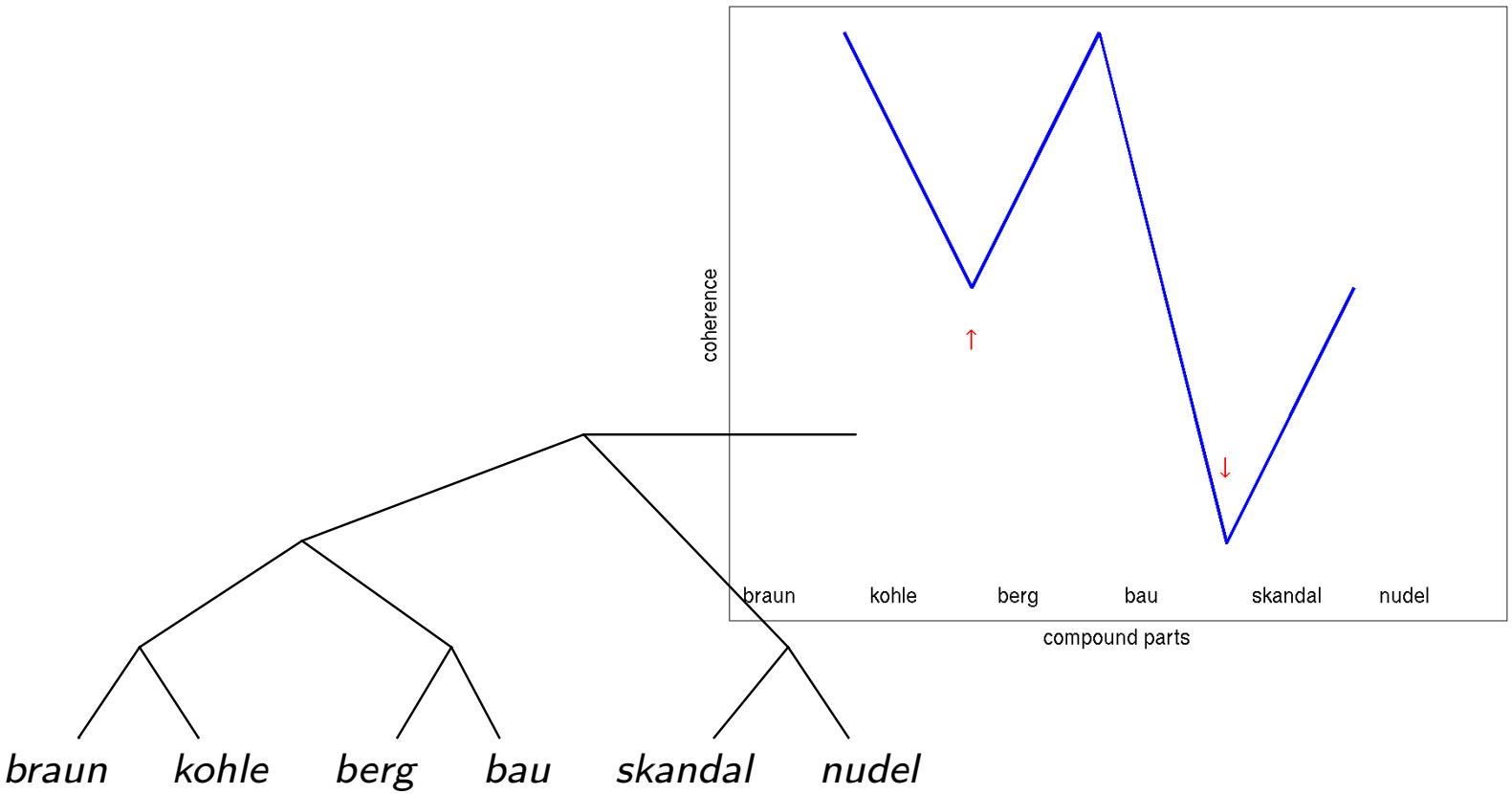
- **Kontrastakzent:**
  - *'Arbeitgeber* **vs.** *Arbeitg'eber und Arbeitn'ehmer*
  - *R'egel* **vs.** *ich habe Reg'el gesagt, nicht Regal*
- **unklare Fälle:** *m'erkwürdigerweise* **vs.** *merkwürdigerw'eise*

Wortbetonung

Vorhersage

1. Kompositumzerlegung (mittels morphologischer Analyse, s.o.)
2. Bestimmung des betonten Kompositumglieds (mittels metrischer Bäume)
3. Lokalisierung der betonten Silbe im betonten Kompositumglied (mittels morphologischer Analyse zur Identifizierung betonter Affixe und maschinellem Lernen)

Betontes Kompositumglied: Induktion metrischer Bäume



Kohärenz: z.B. Bigramm-Wahrscheinlichkeiten: $P(\text{skandal}|\text{bau}) < P(\text{bau}|\text{berg})$

Wortbetonung: Lokalisierung der betonten Silbe

Mittels maschinellem Lernen

- **Objekte:** Silben oder Wörter (Kompositumglieder)
- **Zielwerte:**
  - für Silben: $+/-$ betont
  - für Wörter: Index der betonten Silbe

Wortbetonung: Lokalisierung der betonten Silbe

Instanzbasiertes Lernen (Mustervergleich; Daelemans et al., 1994)

- **Training:** Abspeichern von Wörtern in Form von Merkmalsvektoren zusammen mit Index der betonten Silbe
- **Merkmale:** z.B. Silbengewicht "**s**chwer, **l**eicht" der letzten beiden Silben; **Zielwerte: u**ltima, **p**enultima, **a**ntepenultima

Wort	Merkmalsvektor	Betonung
Wanne	s l	p
Sonne	s l	p
Hibiskus	s s	p
genau	l s	u

Wortbetonung: Lokalisierung der betonten Silbe

- **Anwendung** – für Wort *w*: Übernahme des häufigsten Betonungsmusters unter den *w*-ähnlichsten Wörtern
- **Hamming-Distanz d** zweier Merkmalsvektoren: Anzahl der unterschiedlichen Werte

  w = Wonne (Merkmalsvektor: [s l])
  d(Wonne,Wanne)=0, d(Wonne,Sonne)=0,
  d(Wonne,Hibiskus)=1, d(Wonne,genau)=2
  $\longrightarrow$ Übernahme der Betonung der ähnlichsten Wörter *Wanne*, *Sonne*, also *penultima*

Wortbetonung: Lokalisierung der betonten Silbe

Entscheidungsbaum

- Objekte: Silben
- Features: Silbengewicht, Morphemklasse, POS, Silbenindex, Position in Kompositum
- Zielwert: +/– betont

Evaluierung

- anhand eines Testcorpus mit manueller Transkription (*gold standard*)
- **Word error rate**: Anteil der Wörter, in denen (mindestens) eine Abweichung vom *gold standard* auftritt
- **Phone error rate**: ermittelt über die **Levenshtein-Distanz** zwischen gewünschter (*gold standard*) und berechneter Transkription
- sehr stark abhängig von Sprache / Testkorpus!