

# Abfragen aus einer Sprachdatenbank

Jonathan Harrington

# Aufbau & Funktion von Sprachdatenbanken

Erstellung

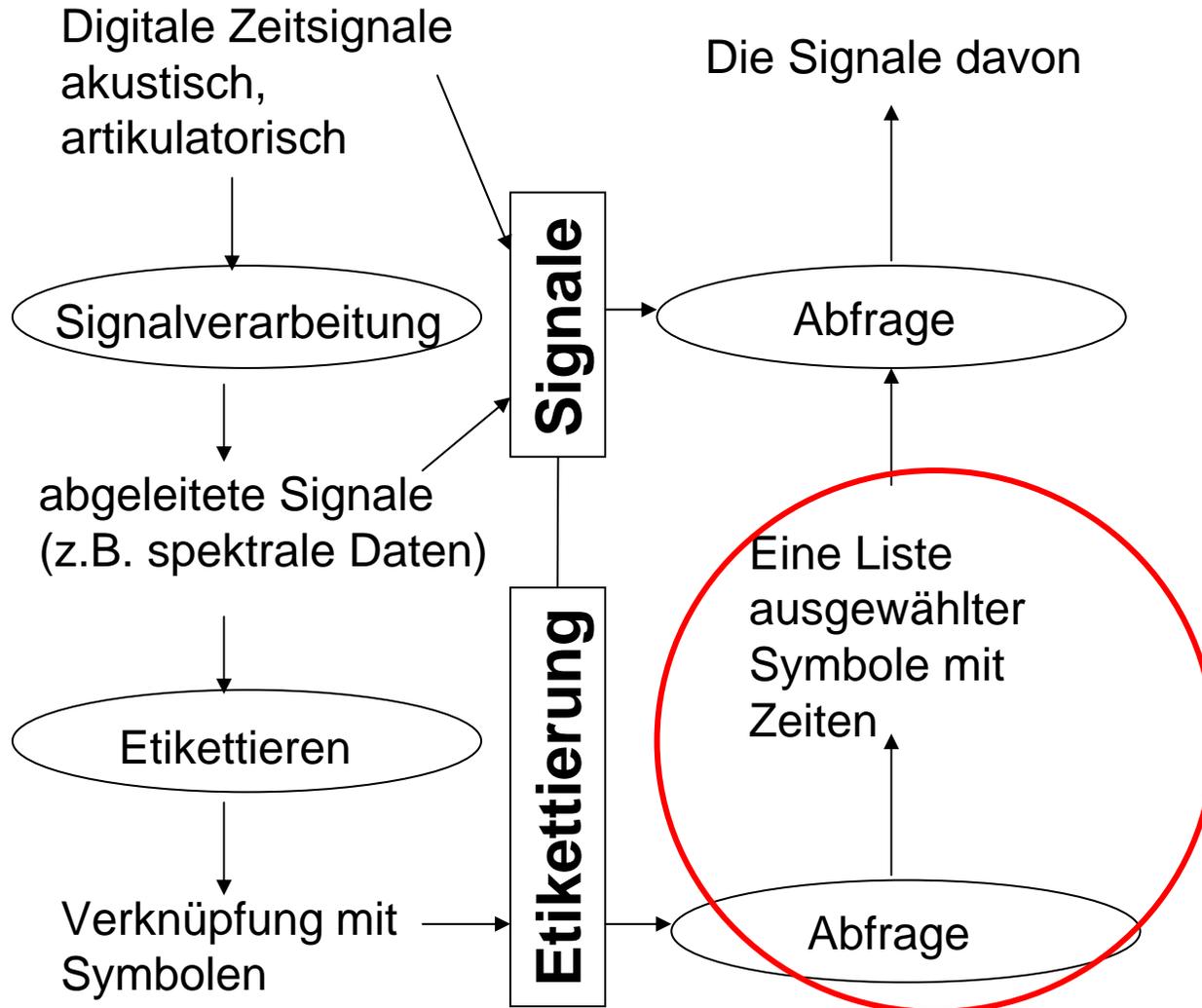
Praat, EMU

Abfrage

EMU

Analyse

EMU-R



# Der Kern einer Abfrage

Welche Ebene?

Welche Einheiten?

Phonetik

u:

(Alle 'u:' Einheiten der **Phonetik** Ebene)

Wort

Duden

(Alle 'Duden' Einheiten der **Wort** Ebene)

Welche Ebene?

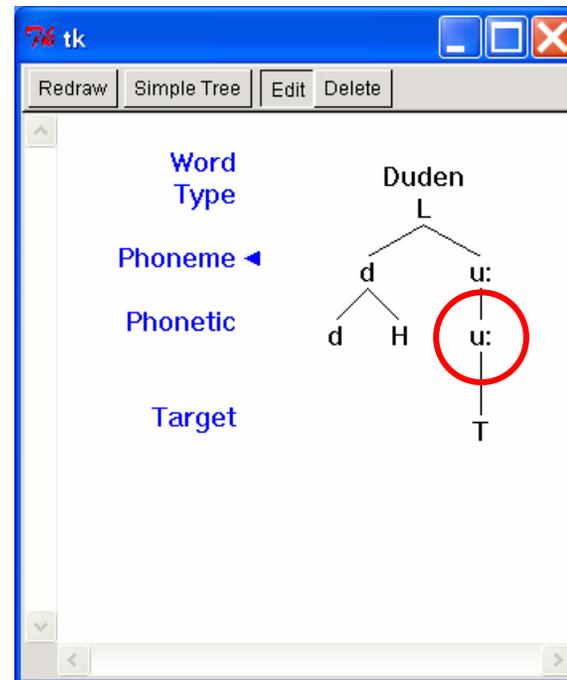
Phonetic

Welche Einheiten?

u:

Datenbank seminar04

Findet



Nicht nur in dieser Äußerung sondern in der gesamten Sprachdatenbank...

# Die Ausgabe einer Abfrage ist eine **Segmentliste**

Label	Start	End	Utterance
u:	509.047	628.372	agr001
u:	566.191	748.833	agr002
u:	536.321	696.716	agr004
u:	539.483	654.833	agr008
u:	586.744	758.656	agr025
u:	531.679	708.31	agr044
u:	560.483	744.803	agr052
u:	509.817	631.886	agr053
u:	512.266	681.855	agr058

Etikettierung

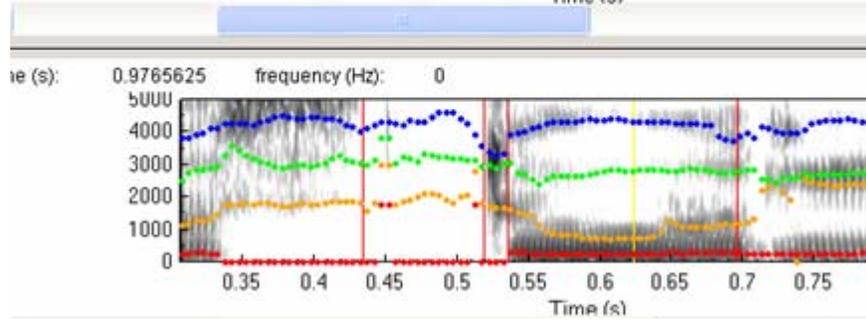
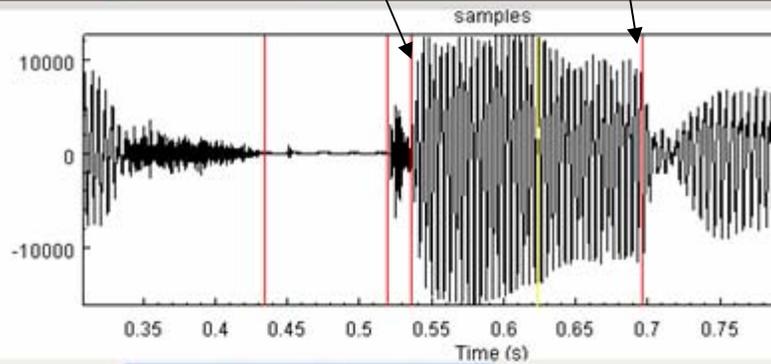
Startzeit

Endzeit

Äußerung

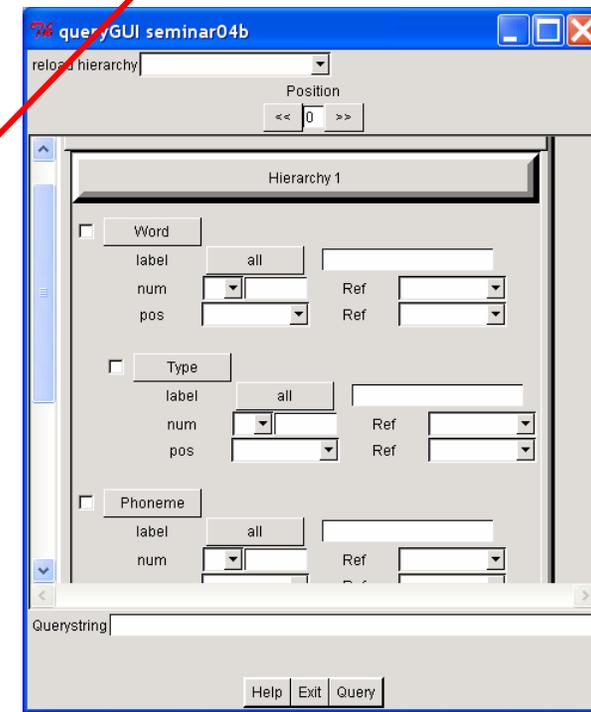
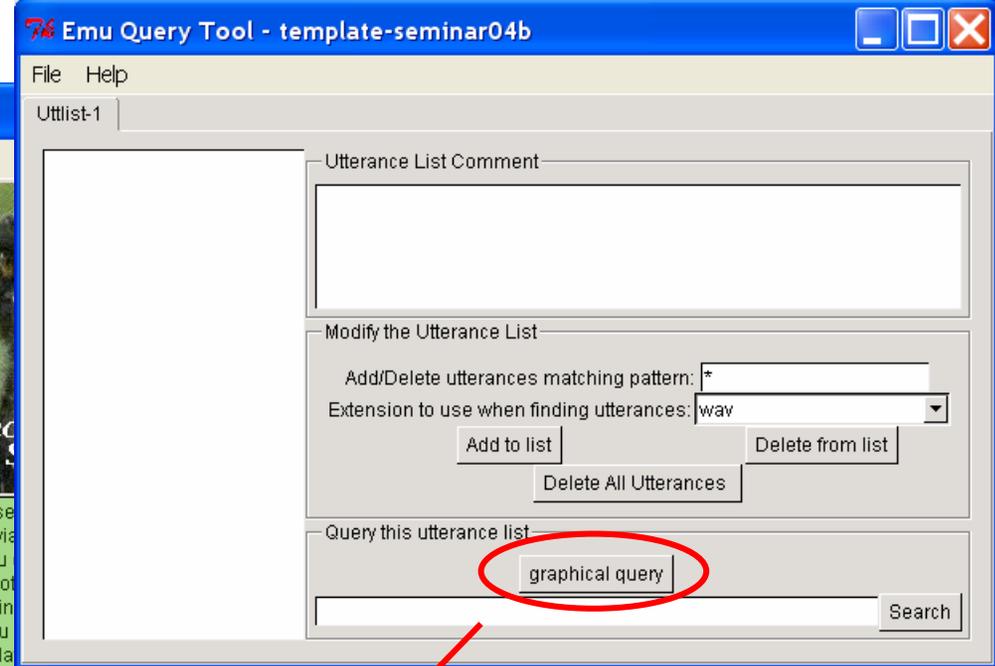
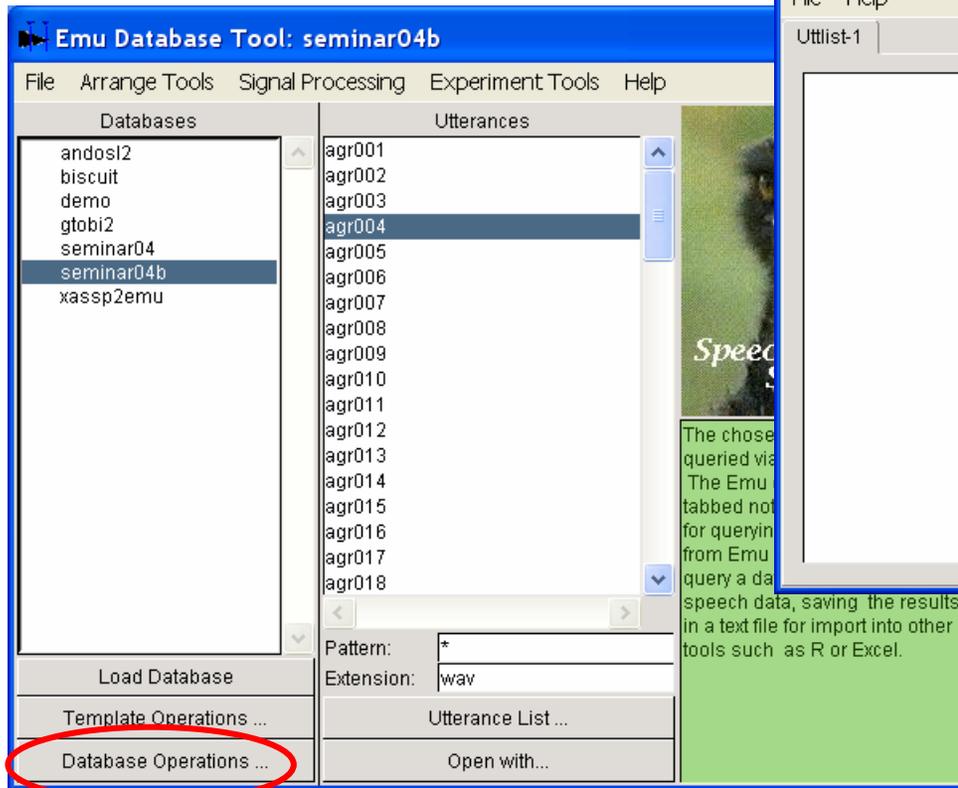
(Die obige Segmentliste enthält alle [u:] der Sprachdatenbank)

Label	Start	End	Utterance
U:	509.047	628.372	agr001
U:	566.191	748.833	agr002
U:	536.321	696.716	agr004
U:	539.192	654.922	agr008
seminar04b/agr004			
Extra Modules Help			
Sample Tree Hierarchy Signals Edit Delete Query			
Phonetic			
d H u:			
Target T			



Beziehung:  
Segmentliste  
und Signale

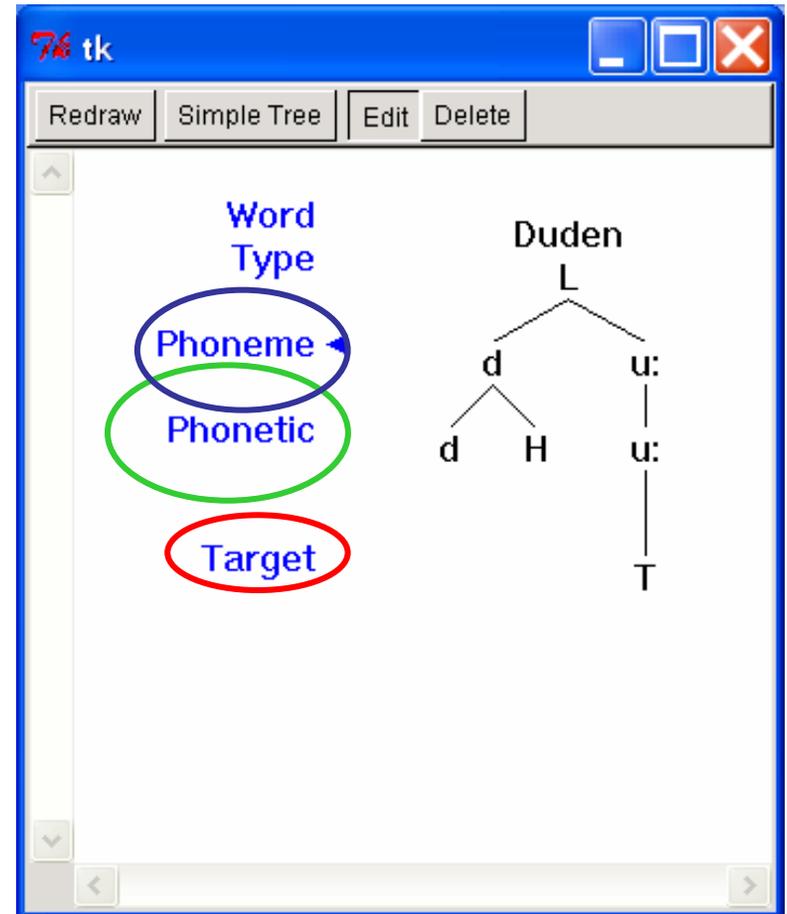
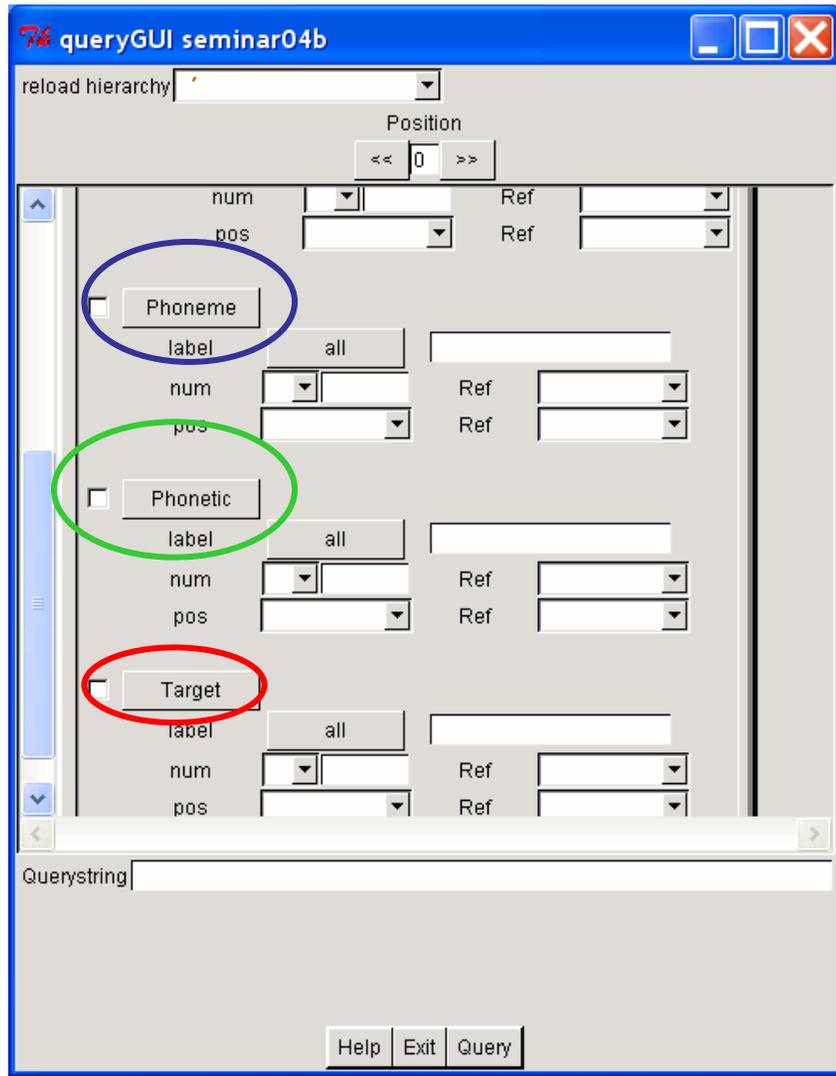
Wie wird eine Abfrage durchgeführt?



1. Query database

2. Graphical query

# Beziehung zwischen Graphical Query und Etikettierungen



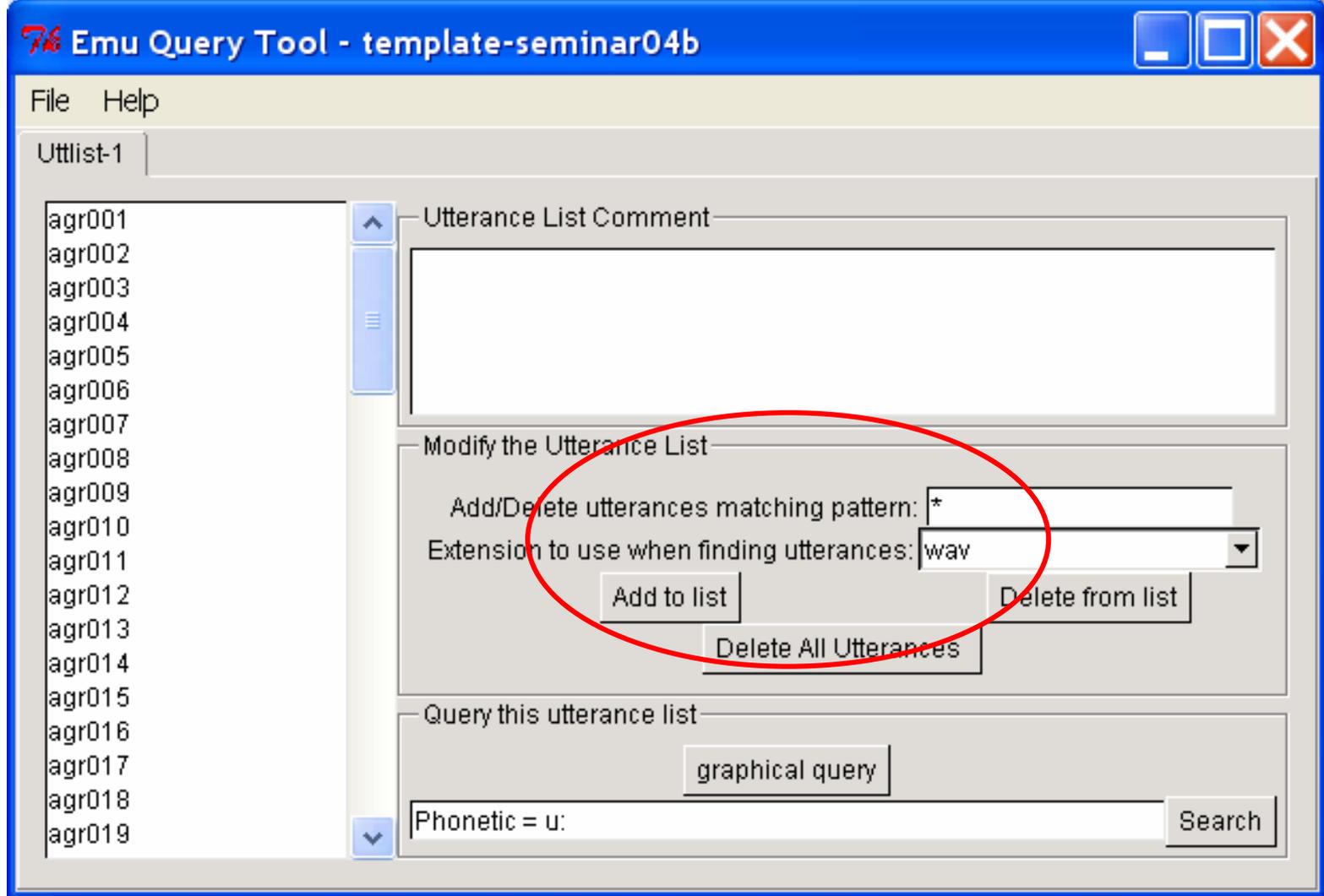
# Phonetic

u:

(Alle 'u:' Einheiten der Phonetic Ebene)

A screenshot of a software interface with a light gray background. On the left, there is a vertical sidebar containing a checked checkbox next to the label 'Phonetic'. To the right of this sidebar, there are several input fields: a text field containing 'all', a text field containing 'u:', and two rows of 'Ref' labels followed by dropdown menus. A red circle highlights the 'Phonetic' checkbox and its label. Another red circle highlights the 'u:' text field.

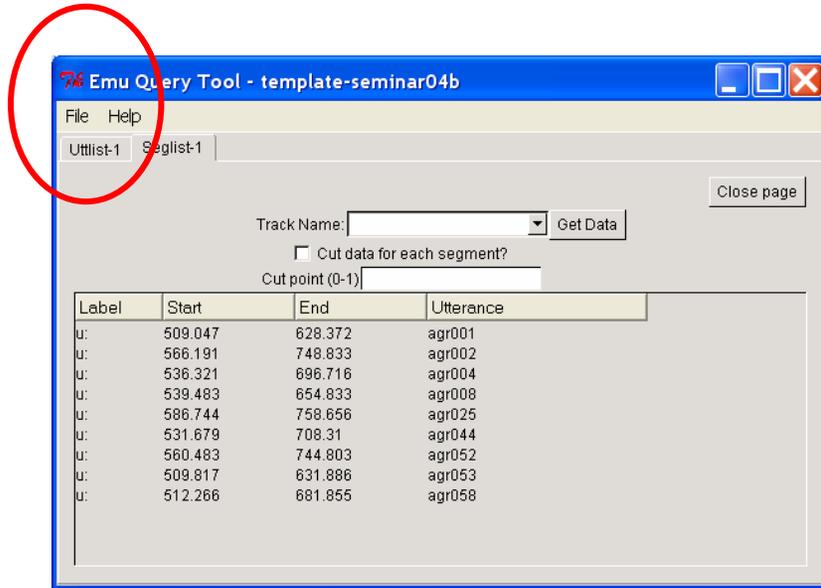
A screenshot of a software interface showing three buttons: 'Help', 'Exit', and 'Query'. A blue circle highlights the 'Exit' and 'Query' buttons. A thick blue horizontal line is positioned below the buttons.



\* wav bedeutet: alle Äußerungen mit dieser Extension abfragen

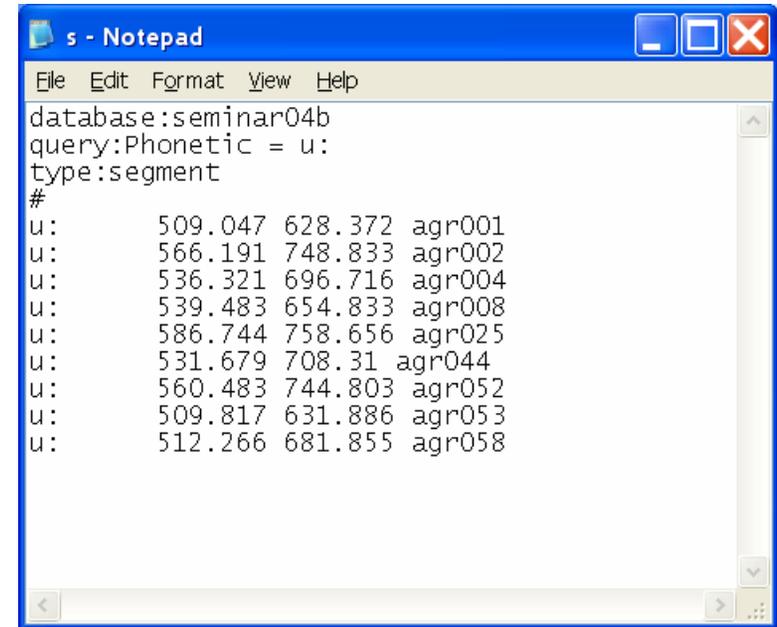
\*01\* lab wäre: alle Äußerungen mit Extension .lab abfragen, die 01 in der Dateiname enthalten

# Ergebnis: Segmentliste



The screenshot shows the 'Emu Query Tool - template-seminar04b' window. The 'Seglist-1' tab is active. The interface includes a 'Track Name' dropdown menu, a 'Get Data' button, a checkbox for 'Cut data for each segment?', and a 'Cut point (0-1)' input field. Below these controls is a table with the following data:

Label	Start	End	Utterance
u:	509.047	628.372	agr001
u:	566.191	748.833	agr002
u:	536.321	696.716	agr004
u:	539.483	654.833	agr008
u:	586.744	758.656	agr025
u:	531.679	708.31	agr044
u:	560.483	744.803	agr052
u:	509.817	631.886	agr053
u:	512.266	681.855	agr058



The screenshot shows a Notepad window with the following text content:

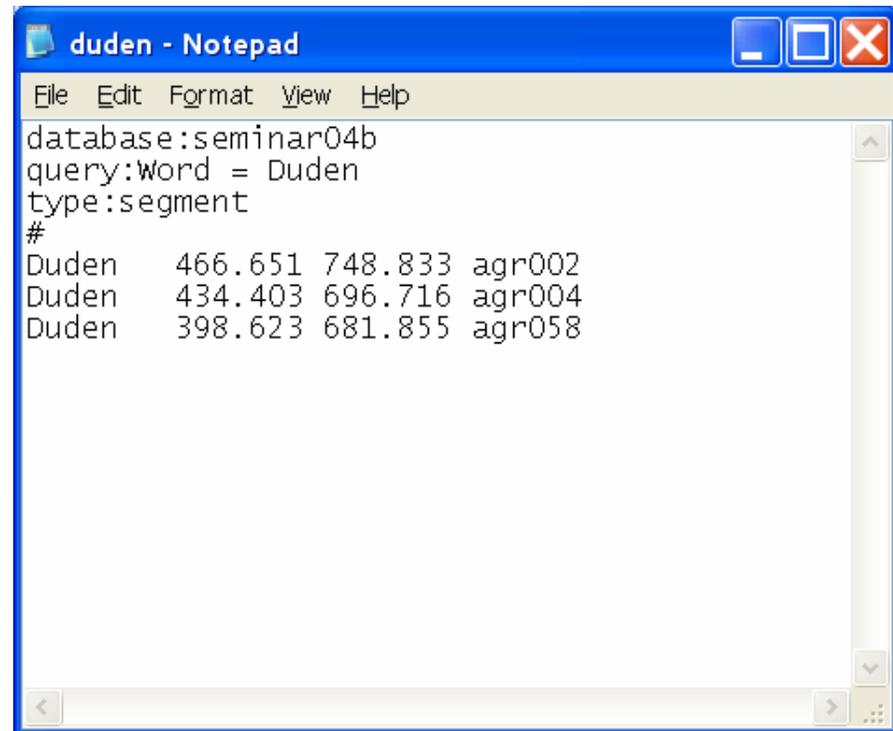
```
database:seminar04b
query:Phonetic = u:
type:segment
#
u:      509.047 628.372 agr001
u:      566.191 748.833 agr002
u:      536.321 696.716 agr004
u:      539.483 654.833 agr008
u:      586.744 758.656 agr025
u:      531.679 708.31  agr044
u:      560.483 744.803 agr052
u:      509.817 631.886 agr053
u:      512.266 681.855 agr058
```

save (Dateiname  
aussuchen)

# Aufgabe

Eine Textdatei aller 'Duden' Einheiten der Word-Ebene speichern

Ergebnis



```
File Edit Format View Help
database:seminar04b
query:Word = Duden
type:segment
#
Duden 466.651 748.833 agr002
Duden 434.403 696.716 agr004
Duden 398.623 681.855 agr058
```

# Weitere Abfragen

## 1. Abfragen aus **einer** Ebene

### 1.1 Kernfragen

**Einfach:**

[u:] Einheiten der **Phonetik** Ebene

**Mehrere Einheiten:**

[u:] und [i:] Einheiten der **Phonetik** Ebene

**Einheiten außer irgendwelchen Einheiten**

Alle Einheiten außer Duden und Diele der **Word** Ebene

### 1.2 Reihenfolge (NB aus **einer** Ebene)

/b/ **Phoneme** vor /i:/ **Phonemen** (also /b/ in /bi:/)

‚Duden‘ der **Word** Ebene gefolgt von ‚ist‘ der **Word** Ebene (also Reihenfolgen von ‚Duden ist‘)

## 2. Abfragen aus mehreren Ebenen

/u:/ **Phoneme** in starken **Silben**

/u:/ und /i:/ **Phoneme** vor /g/ **Phoneme** in starken **Silben**

## 1.1 Kernfragen

### Mehrere Einheiten

A screenshot of a search interface. At the top, a checkbox is checked next to the label 'Phoneme'. Below this, there are three rows of controls: 'label' with a dropdown menu set to 'all', 'num' with a dropdown menu and an empty text box, and 'pos' with a dropdown menu. To the right of these controls is a text input field containing 'u: i:'. Below the text input are two rows, each labeled 'Ref' followed by a dropdown menu.

### Alle Einheiten außer einer/mehreren Einheiten

A screenshot of a search interface. At the top, a checkbox is checked next to the label 'Word'. Below this, there are three rows of controls: 'label' with a dropdown menu set to 'all except', 'num' with a dropdown menu and an empty text box, and 'pos' with a dropdown menu. To the right of these controls is a text input field containing 'Duden'. Below the text input are two rows, each labeled 'Ref' followed by a dropdown menu.

# 1.2 Reihenfolge (NB aus einer Ebene)

/g i:/ Phoneme

The screenshot displays a software interface for defining phoneme positions in a hierarchy. The interface is split into two panes, both labeled "Hierarchy 1".

At the top, a "Position" control shows "<< 1 >>" circled in red.

The left pane shows the following sections:

- Word**: label "all except", num, pos, Ref.
- Type**: label "all", num, pos, Ref.
- Phoneme**: label "g", num, pos, Ref.

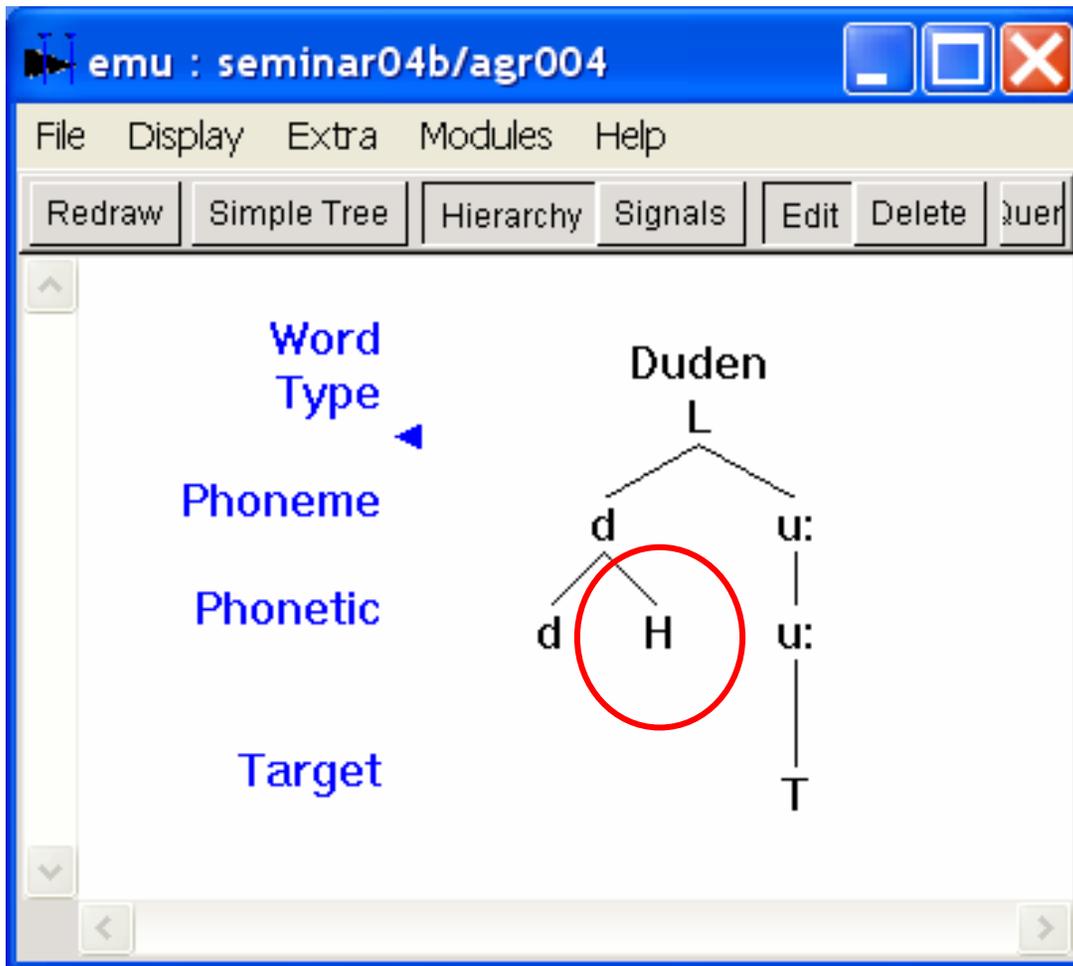
The right pane shows the following sections:

- Word**: label "all", num, pos, Ref.
- Type**: label "all", num, pos, Ref.
- Phoneme**: checked checkbox, label "i:", num, pos, Ref.

Red circles highlight the "Position" control, the "Phoneme" section in the left pane, and the "Phoneme" section in the right pane.

## 2. Abfragen aus mehreren Ebenen

H-Phonetic Einheiten in /d/ Phonemen in L-Einheiten der Type Ebene



# H-Phonetic Einheiten in /d/ Phonemen in L-Einheiten der Type Ebene

Type  
label all L  
num [dropdown] Ref [dropdown]  
pos [dropdown] Ref [dropdown]

Phoneme  
label all d  
num [dropdown] Ref [dropdown]  
pos [dropdown] Ref [dropdown]

Phonetic  
label all H  
num [dropdown] Ref [dropdown]  
pos [dropdown] Ref [dropdown]

## Ergebnis

Cut point (0-1)

Label	Start	End	Utterance
H	543.361	566.191	agr002
H	519.613	536.321	agr004
H	539.407	577.936	agr007
H	582.317	610.776	agr011
H	546.327	567.867	agr015
H	541.608	567.399	agr016
H	538.06	555.417	agr024
H	472.596	498.142	agr027
H	468.978	485.97	agr028
H	503.692	527.331	agr031
H	520.104	540.908	agr033
H	515.24	533.576	agr034
H	457.708	489.584	agr040

# H Einheiten in /d/ Phonemen in Wörtern von Type L

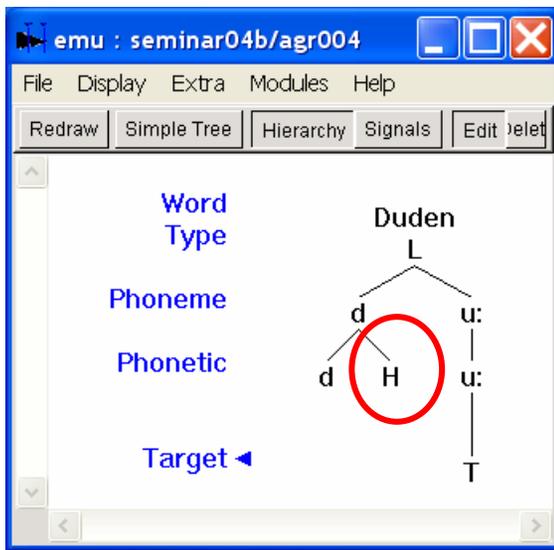
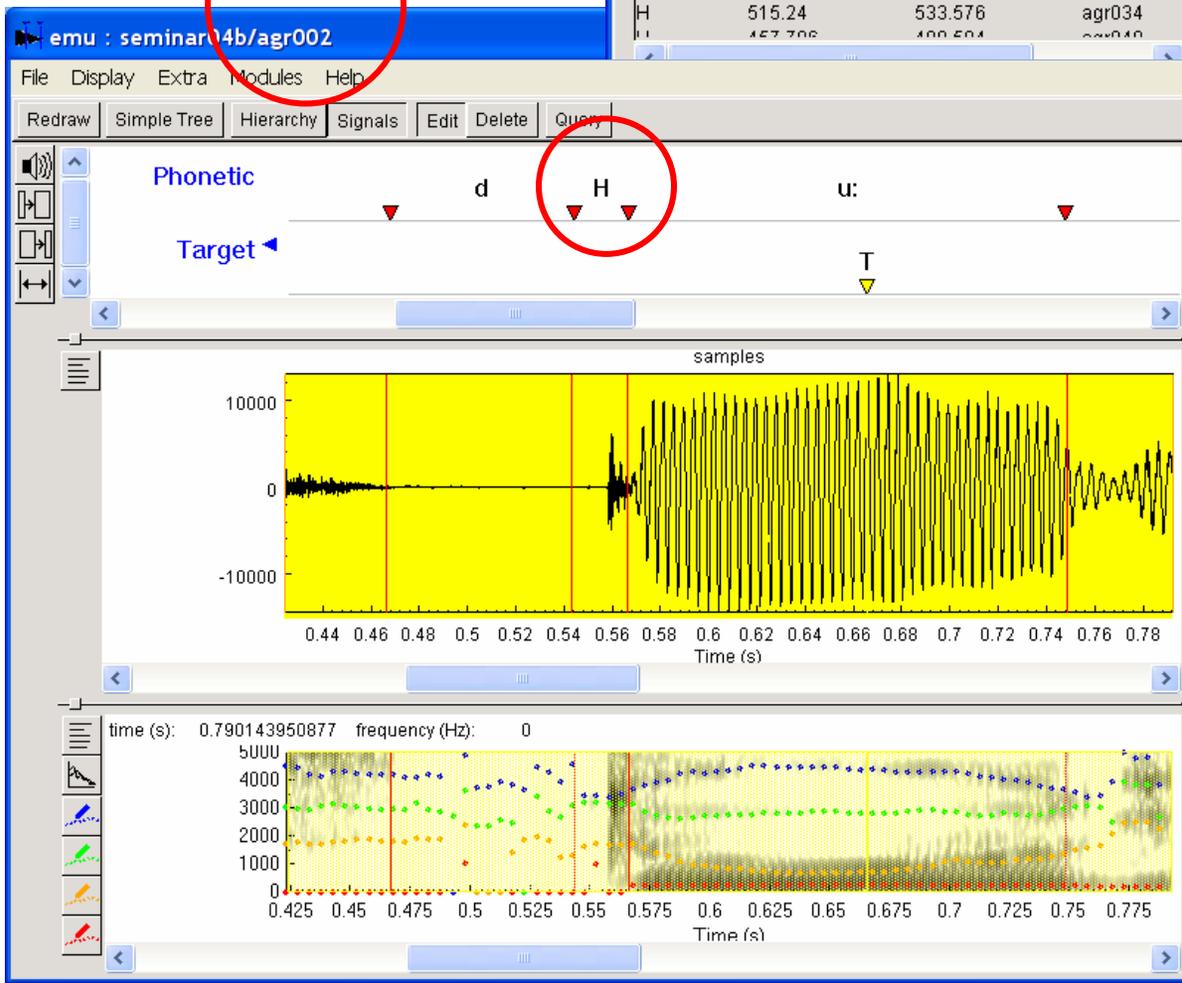
Type  
 label: all  
 num: [ ] Ref: [ ]  
 pos: [ ] Ref: [ ]

Phoneme  
 label: all  
 num: [ ] Ref: [ ]  
 pos: [ ] Ref: [ ]

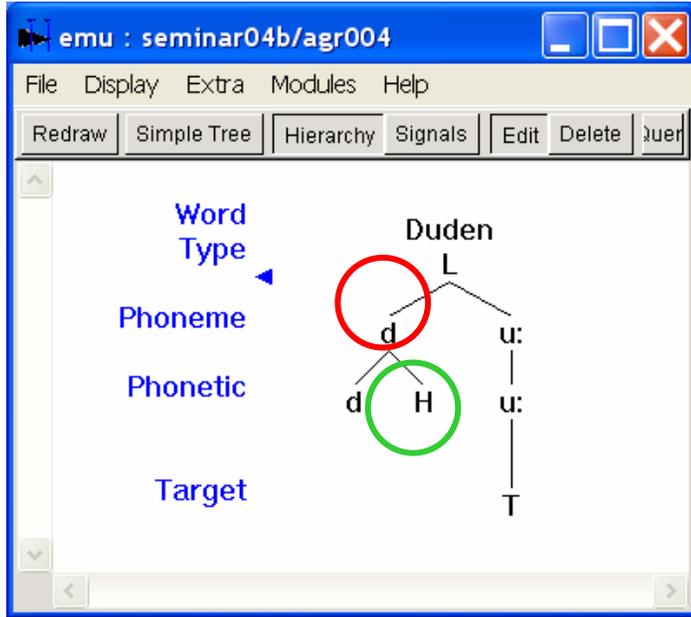
Phonetic  
 label: all  
 num: [ ] Ref: [ ]  
 pos: [ ] Ref: [ ]

Cut point (0-1)

Label	Start	End	Utterance
H	543.361	566.191	agr002
H	519.613	536.321	agr004
H	539.407	577.936	agr007
H	582.317	610.776	agr011
H	546.327	567.867	agr015
H	541.608	567.399	agr016
H	538.06	555.417	agr024
H	472.596	498.142	agr027
H	468.978	485.97	agr028
H	503.692	527.331	agr031
H	520.104	540.908	agr033
H	515.24	533.576	agr034
H	467.706	489.504	agr040



# /d/ Phoneme in Wörtern von Type L, die H Phonetic Einheiten dominieren



Type  
label all L  
num  
pos Ref  
Ref

Phoneme  
label all d  
num Ref  
pos Ref

Phonetic  
label all H  
num Ref  
pos Ref

## Phoneme Einheiten

Label	Start	End	Utterance
d	466.651	566.191	agr002
d	434.403	536.321	agr004
d	479.687	577.936	agr007
d	501.318	610.776	agr011
d	481.708	567.867	agr015
d	458.709	567.399	agr016
d	437.776	555.417	agr024
d	390.485	498.142	agr027
d	397.612	485.97	agr028
d	420.046	527.331	agr031
d	414.192	540.908	agr033
d	445.563	533.576	agr034
d	394.574	480.504	agr040
d	418.436	524.881	agr043

# /d/ Phoneme in Wörtern von Type L, die H Phonetic Einheiten dominieren

Label	Start	End	Utterance
d	466.651	566.191	agr002
d	434.403	536.321	agr004
d	479.687	577.936	agr007
d	501.318	610.776	agr011
d	404.700	507.087	agr015

Type  
 label: all  
 num: [dropdown]  
 pos: [dropdown]

Phoneme  
 label: all  
 num: [dropdown]  
 pos: [dropdown]

Phonetic  
 label: all  
 num: [dropdown]  
 pos: [dropdown]

emu : seminar04b/agr002

File Display Extra Modules Help

Redraw Simple Tree Hierarchy Signals Edit Delete Query

Phonetic: d H u:

Target: T

samples

Time (s): 0.44 0.46 0.48 0.5 0.52 0.54 0.56 0.58 0.6 0.62 0.64 0.66 0.68 0.7 0.72 0.74 0.76 0.78

time (s): 0.790143950877 frequency (Hz): 0

Time (s): 0.425 0.45 0.475 0.5 0.525 0.55 0.575 0.6 0.625 0.65 0.675 0.7 0.725 0.75 0.775

emu : seminar04b/agr004

File Display Extra Modules Help

Redraw Simple Tree Hierarchy Signals Edit Delete

Word Type

Phoneme

Phonetic

Target

Duden

```

  graph TD
    L[L] --- d1[d]
    L --- H[H]
    L --- u1[u:]
    d1 --- d2[d]
    H --- H2[H]
    u1 --- u2[u:]
    u2 --- T[T]
  
```

The phoneme 'd' is circled in red in the diagram.

Fragen: siehe

<http://www.phonetik.uni-muenchen.de/~jmh/>

Lehre > Seminar EMU R > Query Übung