

# EMU : Automatisches Etikettieren – Anwendung von AutoBuild

```

autobuild_default.tcl - Editor
Datei Bearbeiten Format Ansicht ?

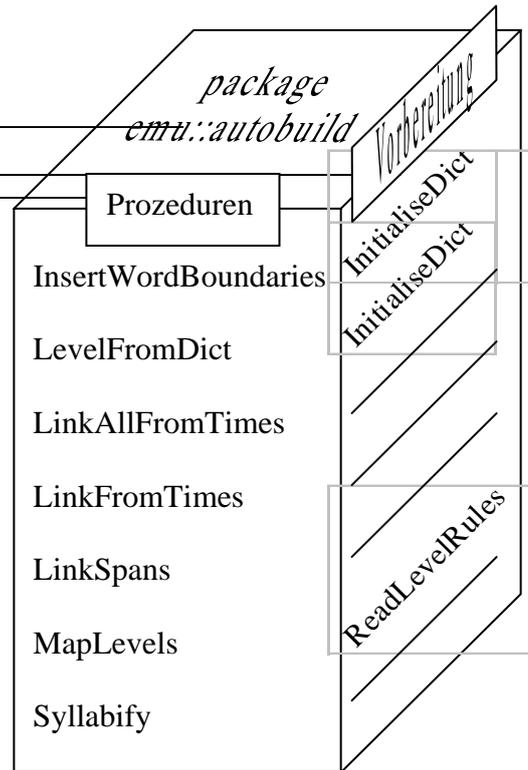
package require emu::autobuild ←

#####
# AutoBuildInit initialises any rulesets
# needed by AutoBuild

proc AutoBuildInit {templ} {
    ←
}

#####
# AutoBuild builds the hierarchy using
# whatever means are appropriate

proc AutoBuild {templ hier} {
    ←
}
    
```



GTemplate Editor w6bsp.tpl

AutoBuild: examples/autobuild\_default.tcl

HierarchyViewLevels:

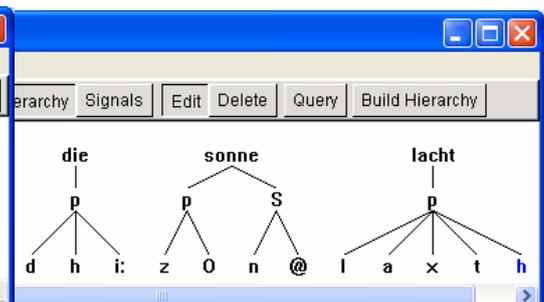
- 1|Word
- 2|Syllable
- 3|Phoneme

emu : w6bsp/K01BE002

Word \* die sonne lacht \*

Syllable \* p p S p \*

Phoneme \* d h i: z O n @ la x t h \*



# Übung

## Ziel:

- Unter der Verwendung von AutoBuild-Scripts sollen Äußerungen einer Sprachdatenbank automatisch etikettiert werden

## Aufgabe I:

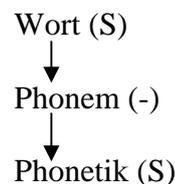
In *Aufgabe III* des Übungspapers mit dem Thema: *EMU: Erstellung einer Templatedatei mit Links* (EMU\_GTed\_LinksUebung.pdf) sollten 2 der 5 *gam* Äußerungen Links zwischen den Segmenten auf den drei verschiedenen Ebenen hinzugefügt werden.

Füge nun diese Links für die restlichen 3 *gam* Äußerungen automatisch hinzu. Hierfür:

- Schreibe ein AutoBuild-Script, in dem die Prozedur **LinkAllFromTimes** aufgerufen wird. Hierfür:
  - Öffne einen Editor: Xemacs oder nedit oder wie du willst
  - Schreibe die erforderlichen Zeilen
  - Speicher diese Datei unter einem beliebigen Namen irgendwohin
- Füge in deiner Template-Datei zu den *gam* Äußerungen im *GTed* auf der Registerkarte *Variables* im Eingabefeld *Autobuild* den Pfad zu dem geschriebenen AutoBuild-Script ein.
- Öffne die *gam* Äußerungen in *emulabel* und klicke auf den Button *Build Hierarchy*.

## Aufgabe II:

1. **Etikettiere** die Äußerung *vormittag* im Verzeichnis *dbs/autobuildI* unter */data/teach/Modul\_F* auf den Ebenen Wort und Phonetik der Kette (verlinke keine Segmente!):



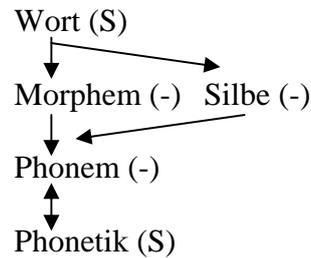
2. Lass die Phoneme auf der Phonem-Ebene **automatisch erstellen**. Hierfür:
  - Schreibe ein AutoBuild\_Script, in dem die Prozedur **MapLevels** aufgerufen wird.
    - MapLevels benötigt als Vorbereitung ein Regelwerk, welches Phon zu welchem Phonem gehört
    - Schreibe das Regelwerk in eine Textdatei und speichere sie unter beliebigen Namen.
      - Wir beschränken uns auf die Regel: t h -> t
    - Füge der Vorbereitungsprozedur des AutoBuild-Scriptes *AutoBuildInit* den Prozeduraufruf *ReadLevelRules*, mit dem das Regelwerk verarbeitet werden kann hinzu
    - Vergiss die weiteren benötigten Angaben nicht – orientiere dich an der Vorlesung (S. 16)
3. Lass die Segmente auf den Ebenen Wort und Phonetik automatisch verbinden.

## AutoBuild-Prozeduren im package emu::autobuild

In proc AutoBuild {templ hier} {...}	In proc AutoBuildInit {templ} {...}	Form der zusätzlichen txt Datei
<b>InsertWordBoundaries</b> \$templ \$hier dictionary plevel clevel	InitialiseDict dictionary dict.txt	die d *i: sonne s *O n @ lacht l *a x t
<b>LevelFromDict</b> \$hier plevel clevel dictionary	InitialiseDict dictionary dict.txt	die d *i: sonne s *O n @ lacht l *a x t
<b>LinkAllFromTimes</b> \$templ \$hier		
<b>LinkFromTimes</b> \$hier plevel clevel		
<b>LinkSpans</b> \$hier plevel clevel		
<b>MapLevels</b> \$templ \$hier plevel clevel \$rules	set rules [ReadLevelRules rules.txt]	<s=stop> h -> <s> t s -> ts p f -> pf
<b>Syllabify</b> \$templ \$hier clevel plevel cons	<b>Vor Syllabify</b> Aufruf in die Prozedur <b>AutoBuild</b>  set cons(triples) {pfl pfr tsv Spr Spl Str skr skl }  set cons(pairs) {pr br tr dr kr gr fr Sr pl bl kl gl fl Sl Sm kn gn Sn kv Sv Sp St sk}  set cons(singles) {p b t d k g pf ts dZ f v s z S Z C x m n N l r j h}	In der <b>Template-Datei</b> müssen legal labels für vowel (alle mgl. Nukleus Vokale) angegeben sein

### Aufgabe III (ALS KLAUSURVORBEREITUNG – HAUSAUFGABE!):

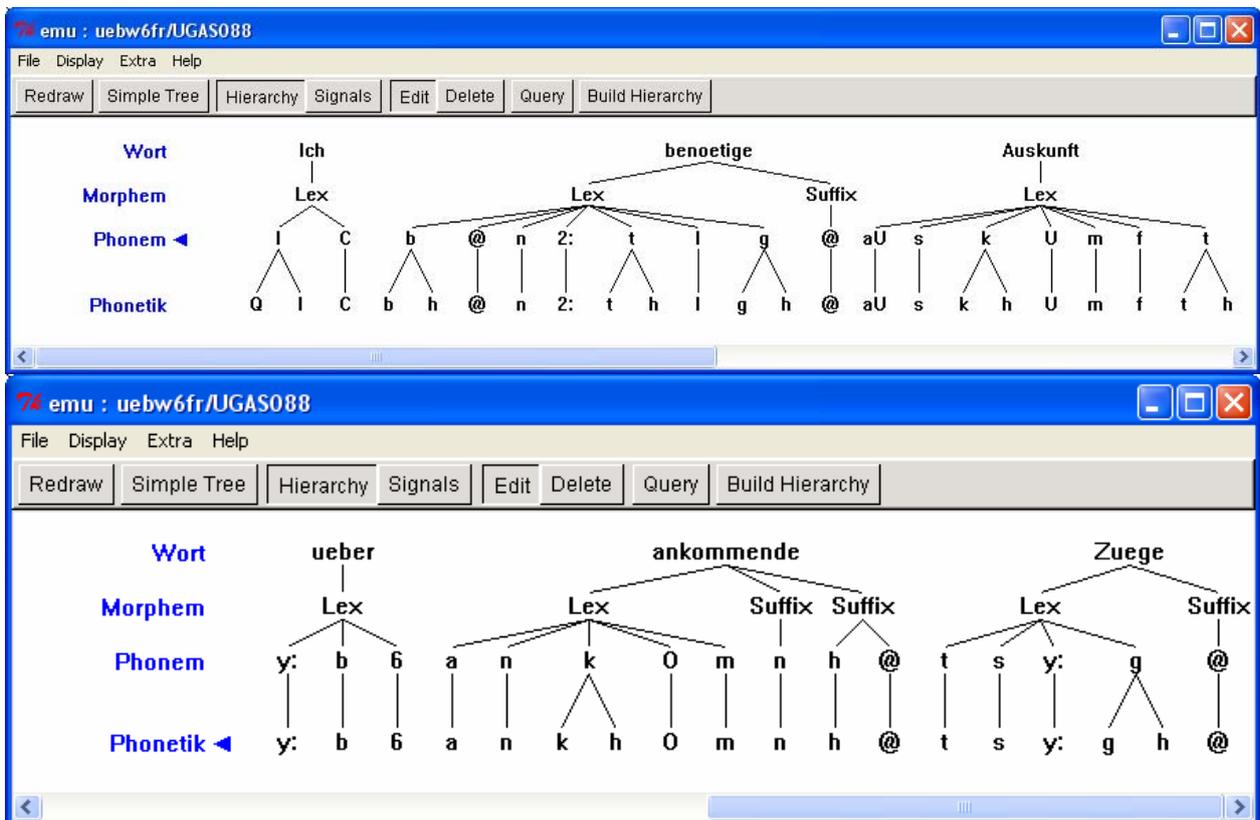
Etikettiere die Äußerung *UGAS088* im Verzeichnis *db/autobuildII* unter */data/teach/Modul\_F* **automatisch** auf den Ebenen der Kette:



Die Ebenen Wort und Phonetik sind bereits etikettiert. Die Etikettierungsdateien *UGAS088.Phonetik* und *UGAS088.Wort* befinden sich im selben Verzeichnis, wie die Äußerung selbst (*UGAS088.wav*). Kopiere dir die beiden Etikettierungsdateien in ein Verzeichnis, in das du deine Etikettierungsdateien für diese Aufgabe speichern möchtest.

1. Erzeuge auf Grundlage der Phonetik-Ebene die Segmente auf der Phonem-Ebene<sup>1</sup>.
2. Erzeuge auf Grundlage der Wort-Ebene die Segmente auf der Morphem-Ebene<sup>2</sup>.
  - Auf der Morphem-Ebene sollen Lex und Suffix erscheinen
3. Lasse nun die Segmente auf der Morphem-Ebene mit den Segmenten auf der Phonem-Ebene verbinden.<sup>3</sup> ACHTUNG: hier müssen vorher manuell ein paar Links gesetzt werden.

Zusatz: Erzeuge auf Grundlage der Phonem-Ebene die Segmente auf der Silben-Ebene.



<sup>1</sup> 1. und 2. können in einem AutoBuild-Script kombiniert werden.

<sup>2</sup> Nachdem 1. und 2. funktionieren, speichere die Äußerungen, so wie sie nach den AutoBuilds aussieht-Vorbereitung zu 3.

<sup>3</sup> Hier wird ein neues AutoBuild-Script benötigt.