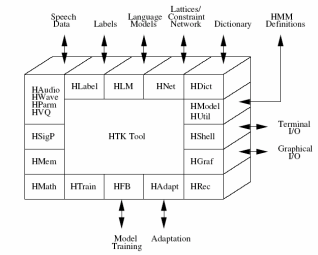
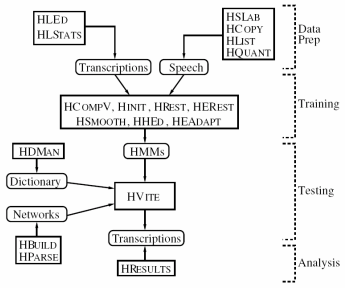
Biblioteka HTK

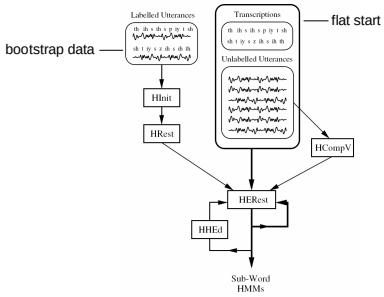
Biblioteka HTK



Etapy przetwarzania danych z wykorzystaniem modułów HTK



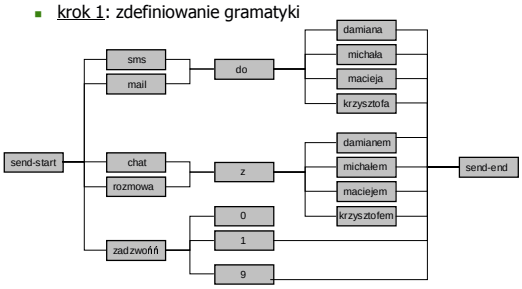
Etap uczenia



Tworzenie systemu rozpoznawania mowy z wykorzystaniem biblioteki **HTK**

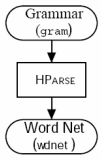
Przygotowanie danych

- krok 1: zdefiniowanie gramatyki



Przykład pliku typu *gram*

```
$liczby = jeden | dwa | trzy | cztery | pięć | sześć | siedem | osiem | dziewięć | zero;
$imie_narzednik = damianem | michałem | maciejem | krzysztofem;
$imie_dopelniacz = damiana | michała | macieja | krzysztofa;
$polecenia1 = sms | mail;
$polecenia2 = chat | rozmowa;
(send-start ($polecenia1 do $imie_dopelniacz | $polecenia2 z $imie_narzednik) | (zadzwoń <$liczby>) send-end)
```

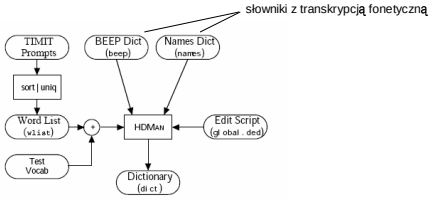


przykład pliku typu *wdnet*

```
VERSION=1.0
N=33  L=61
I=0   W=send-end
I=1   W=!NULL
I=2   W=zero
I=3   W=!NULL
I=4   W=dziew\352\346
I=5   W=osiem
I=6   W=siedem
I=7   W=sze\234\346
I=8   W=pi\352\346
I=9   W=cztery
I=10  W=trzy
I=11  W=dwa
I=12  W=jeden
I=13  W=zadzwoń
I=14  W=krzysztofem
I=15  W=maciejem
I=16  W=micha\263em
I=17  W=damianem
I=18  W=z
I=19  W=rozmowa
I=20  W=!NULL
I=21  W=chat
I=22  W=krzysztofa
I=23  W=macieja
I=24  W=micha\263a
I=25  W=damiana
I=26  W=do
I=27  W=mail
I=28  W=!NULL
I=29  W=sms
I=30  W=send-start
I=31  W=!NULL
I=32  W=!NULL
J=0   S=3   E=0
J=1   S=0   E=1
J=2   S=14  E=1
J=3   S=15  E=1
J=4   S=16  E=1
J=5   S=17  E=1
J=6   S=22  E=1
J=7   S=23  E=1
J=8   S=24  E=1
J=9   S=25  E=1
J=10  S=3   E=2
J=11  S=13  E=2
J=12  S=2   E=3
J=13  S=4   E=3
J=14  S=5   E=3
J=15  S=6   E=3
J=16  S=7   E=3
J=17  S=8   E=3
J=18  S=9   E=3
J=19  S=10  E=3
J=20  S=11  E=3
J=21  S=12  E=3
J=22  S=3   E=4
J=23  S=13  E=4
J=24  S=3   E=5
J=25  S=13  E=5
J=26  S=3   E=6
J=27  S=13  E=6
J=28  S=3   E=7
J=29  S=13  E=7
J=30  S=3   E=8
J=31  S=13  E=8
J=32  S=3   E=9
J=33  S=13  E=9
J=34  S=3   E=10
J=35  S=13  E=10
J=36  S=3   E=11
J=37  S=13  E=11
J=38  S=3   E=12
J=39  S=13  E=12
J=40  S=32  E=13
J=41  S=18  E=14
J=42  S=18  E=15
J=43  S=18  E=16
J=44  S=18  E=17
J=45  S=20  E=18
J=46  S=30  E=19
J=47  S=19  E=20
J=48  S=21  E=20
J=49  S=30  E=21
J=50  S=26  E=22
J=51  S=26  E=23
J=52  S=26  E=24
J=53  S=26  E=25
J=54  S=28  E=26
J=55  S=30  E=27
J=56  S=27  E=28
J=57  S=29  E=28
J=58  S=30  E=29
J=59  S=32  E=30
J=60  S=1   E=31
```

Moduł *HDMan*

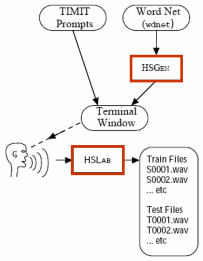
- krok 2: słownik z transkrypcją fonetyczną
  - utworzenie posortowanej listy słów – *wlist*
  - zastosowanie modułu *HDMan* do wygenerowania słownika



Przykładowy plik *dict*

jeden	j e d e n sp
dwa	d v a sp
trzy	t s6 y sp
cztery	c6 t e r y sp
pięć	p j e ń ć sp
sześć	s6 e ś ć sp
siedem	ś e d e m sp
osiem	o ś e m sp
dziewięć	Z@ e v j e ń Z@ sp
zero	z e r o sp
zadzwońń	z a Z v o ń sp
wywołaj	v y v o w a j sp
damiana	d a m j a n a sp
damianem	d a m j a n e m sp
michała	m i X a w a sp
michałem	m i X a w e m sp
macieja	m a ć e j a sp
maciejem	m a ć e j e m sp
krzysztofa	k s6 y s6 t o f a sp
krzysztofem	k s6 y s6 t o f e m sp
chat	c6 a d sp
z	z sp
do	d o sp
rozmowa	r o z m o v a sp
sms	e s e m e z sp
mail	m e j l sp
send-end []	sil
send-start[]	sil

Rejestracja i etykietowanie danych akustycznych



HSGen

- Przygotowanie pliku z *promptami:*

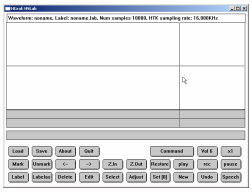
```
HSGen -l -n 18 wdnet dict > testprompts
```

Przykładowy plik *testprompts:*

1. chat z maciejem
2. zadzwoń pięć cztery trzy pięć trzy sześć jeden trzy sześć trzy sześć sześć cztery zero siedem trzy
3. chat z maciejem
4. chat z michałem
5. zadzwoń jeden sześć trzy dwa
6. zadzwoń pięć zero zero pięć trzy osiem dziewięć trzy dziewięć sześć dwa jeden jeden dwa
7. chat z krzysztofem
8. zadzwoń sześć siedem cztery dziewięć dwa dwa jeden cztery pięć zero sześć zero
9. zadzwoń osiem dwa siedem cztery dziewięć dziewięć cztery jeden
10. rozmowa z maciejem
11. chat z maciejem
12. sms do damiana
13. mail do krzysztofa
14. mail do michała
15. zadzwoń osiem
16. zadzwoń pięć trzy dziewięć dwa zero dwa dwa zero
17. zadzwoń siedem siedem pięć dwa dwa osiem sześć jeden trzy zero
18. zadzwoń osiem siedem siedem osiem dwa dziewięć dwa jeden zero sześć

Moduł *HSLab*

- krok 3: rejestracja i oznaczenie nagrań akustycznych



przykład pliku 'prompt batch':

```
import os
import msvcrt

output = ["noname_0", "noname_1"]
path = "C:\\htk\\"
text = file(path+"testprompts", "r").readlines()
id = 1
i = 0
for s in text:
  print s
  c = msvcrt.getch()
  if c == 'q':
    break
  os.rename(path+output[i], path+str(id)+".wav")
  i = (i+1) % 2
  id += 1
```

output:
noname_0
noname_1

prompt batch

1.wav
2.wav
…
n.wav

Transkrypcja

- krok 4: generowanie plików z transkrypcją

1. chat z maciejem
2. zadzwoń pięć cztery trzy pięć trzy sześć jeden trzy sześć trzy sześć sześć cztery zero siedem trzy
3. ...

transkrypcja ortograficzna (*prompts2mlf*)

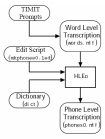
#!MLF!#	sześć
"1.lab"	jeden
chat	trzy
z	sześć
maciejem	
.	trzy
"2.lab"	sześć
zadzwoń	sześć
pięć	cztery
cztery	zero
trzy	siedem
pięć	trzy
trzy	.
	"3.lab„
	...

transkrypcja cd.

words.mlf

#!MLF!#	sześć
"1.lab"	jeden
chat	trzy
z	sześć
maciejem	
.	trzy
"2.lab"	sześć
zadzwoń	sześć
pięć	cztery
cztery	zero
trzy	siedem
pięć	trzy
trzy	.
	"3.lab„
	...

transkrypcja fonetyczna (HLEd)

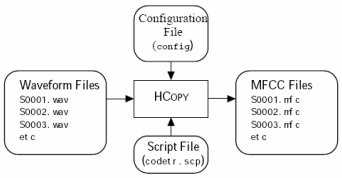


phones0.mlf

#!MLF!#	p	s6	s6	t
"1.lab"	j	e	e	s6
sil	e	\234	\234	y
c6	\361	\346	\346	sil
a	\346	j	s6	"3.lab"
d	c6	e	e	...
z	t	d	\234	
m	e	e	\346	
a	r	n	c6	
\346	y	t	t	
e	t	s6	e	
j	s6	y	r	
e	y	s6	y	
m	p	e	z	
sil	j	\234	e	
.	e	\346	r	
"2.lab"	\361	t	o	
sil	\346	s6	\234	
z	t	y	e	
a	s6		d	
Z	y		e	
v			m	
o				
\361				

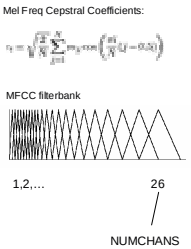
Kodowanie – moduł *HCopy*

- krok 5: ekstrakcja cech akustycznych z plików *wav*



Plik konfiguracyjny (***config***) modułu HCopy

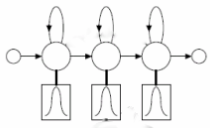
```
TARGETKIND     = MFCC_0
TARGETRATE     = 100000
SAVECOMPRESSED = TRUE
SAVEWITHCRC    = TRUE
WINDOWSIZE     = 250000.0
USEHAMMING     = TRUE
PREEMCOEF      = 0.97
NUMCHANS       = 26
CEPLIFTER      = 22
NUMCEPS        = 12
ENORMALISE     = FALSE
```



Tworzenie modeli HMM dla monofonów

- krok 6: tworzenie prototypowego modelu – ***proto***
  (definiowanie topologii modelu, np. 3-stanowe modele HMM)

```
~o <VecSize> 39 <MFCC_0_D_A> <StreamInfo> 1 39
~h "proto"
<BeginHMM>
  <NumStates> 5
  <State> 2 <NumMixes> 1
  <Stream> 1
  <Mixture> 1 1.0000
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0…
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0…
  <State> 3 <NumMixes> 1
  <Stream> 1
  <Mixture> 1 1.0000
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0…
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0…
  <State> 4 <NumMixes> 1
  <Stream> 1
  <Mixture> 1 1.0000
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0…
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0…
  <TransP> 5
    0.000e+0   1.000e+0   0.000e+0   0.000e+0   0.000e+0
    0.000e+0   6.000e-1   4.000e-1   0.000e+0   0.000e+0
    0.000e+0   0.000e+0   6.000e-1   4.000e-1   0.000e+0
    0.000e+0   0.000e+0   0.000e+0   6.000e-1   4.000e-1
    0.000e+0   0.000e+0   0.000e+0   0.000e+0   0.000e+0
<EndHMM>
```



Moduł *HCompV*

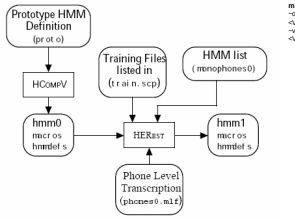
Użycie:
HCompV –C **config** –f 0.01 –m –S train.scp – M **hmm0 proto**

```
~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULL><MFCC_D_A_0><DIAGC>
~h "proto"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 39
 -9.330825e+000 -1.454272e+000 9.844367e+000 -8.358783e+000 3.545721e+000 -2.563990e+000 -3.068573e+000 1.363987e+000 -
1.616646e+000 3.443969e+000 -2.147115e+000 -1.690576e+000 6.517564e+001 -2.458594e-003 1.284420e-003 -1.878561e-004 -9.098755e-004
-3.436189e-004 3.561667e-003 5.398354e-003 -1.636389e-003 -2.713527e-004 -3.456994e-003 -3.376979e-004 -1.380631e-003 -3.194355e-
003 3.325619e-004 -5.367188e-005 -5.679399e-004 -4.837528e-004 2.658523e-005 2.076999e-004 1.000405e-003 1.883085e-004 1.655696e-
004 8.256703e-004 5.778460e-004 7.446198e-004 -2.471439e-004
<VARIANCE> 39
 7.326858e+001 2.338221e+001 4.921546e+001 2.679774e+001 3.580824e+001 2.873628e+001 3.588305e+001 2.879663e+001 2.571888e+001
2.381734e+001 2.283296e+001 1.448675e+001 4.023338e+001 3.235048e+000 1.501695e+000 1.671151e+000 1.825641e+000 2.064622e+000
2.096886e+000 2.637929e+000 1.996072e+000 2.015897e+000 1.671477e+000 1.722597e+000 1.300004e+000 1.555545e+000 4.194194e-001
2.614676e-001 2.610585e-001 3.221001e-001 3.574921e-001 3.841705e-001 4.902699e-001 3.786086e-001 3.804388e-001 3.205249e-001
3.264408e-001 2.545738e-001 2.420380e-001
<GCONST> 1.180158e+002
<STATE> 3
<MEAN> 39
<MEAN> 39
 -9.330825e+000 -1.454272e+000 9.844367e+000 -8.358783e+000 3.545721e+000 -2.563990e+000 -3.068573e+000 1.363987e+000 -
1.616646e+000 3.443969e+000 -2.147115e+000 -1.690576e+000 6.517564e+001 -2.458594e-003 1.284420e-003 -1.878561e-004 -9.098755e-004
-3.436189e-004 3.561667e-003 5.398354e-003 -1.636389e-003 -2.713527e-004 -3.456994e-003 -3.376979e-004 -1.380631e-003 -3.194355e-
003 3.325619e-004 -5.367188e-005 -5.679399e-004 -4.837528e-004 2.658523e-005 2.076999e-004 1.000405e-003 1.883085e-004 1.655696e-
004 8.256703e-004 5.778460e-004 7.446198e-004 -2.471439e-004
<VARIANCE> 39
 7.326858e+001 2.338221e+001 4.921546e+001 2.679774e+001 3.580824e+001 2.873628e+001 3.588305e+001 2.879663e+001 2.571888e+001
~
```

Re-estymacja modeli *hmm0*

pruning parameters:

```
Użycie 3x:
HERest –C config –I phones0.mlf –t 250.0 150.0 1000.0 \
-S train.scp –H hmm0/macros –H hmm0/hmmdefs –M hmm1 monophones0
```

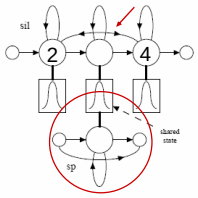


Wynik re-estymacji hmm*

```
~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULL><MFCC_D_A_0><DIAG>
~h "x"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 39
 -1.572051e+001 -1.156194e+000 1.248590e+001 -1.417209e+001 -1.236995e+000 5.387733e+000 -3.952639e+000 -1.244066e+000 -8.989190e-001
5.475657e+000 -2.150294e+000 -1.389184e+000 7.121432e+001 2.520046e-001 5.690539e-001 6.951439e-001 1.330935e+000 1.234478e+000
3.184911e-001 6.900452e-001 5.331070e-001 2.446077e-001 -1.881203e+000 -6.519085e-001 1.457298e-001 -1.285166e+000 1.128686e+000 -
4.766131e-002 1.457025e-002 6.681462e-001 2.875038e-001 -6.105800e-001 2.406946e-001 1.356167e-001 -2.481063e-001 -5.370696e-002 -
1.655559e-002 -2.363150e-002 -4.935381e-001
<VARIANCE> 39
 2.124186e+002 4.119287e+000 4.117410e+001 2.345038e+001 1.440223e+001 1.822623e+001 1.356726e+001 9.836095e+000 1.140151e+001
1.457696e+001 1.248199e+001 5.477507e+000 7.570092e+000 6.931587e+000 1.392687e+000 9.778274e-001 1.749599e+000 1.141136e+000
4.121767e+000 2.196367e+000 1.308992e+000 6.467630e-001 7.021763e-001 6.666625e-001 3.186259e-001 6.070709e-001 1.133426e+000
3.408346e-001 1.331865e-001 4.917689e-001 4.515935e-001 4.664707e-001 4.412114e-001 2.159842e-001 4.795822e-002 1.225621e-001
3.772313e-001 1.032322e-001 4.332458e-001
<GCONST> 9.233756e+001
<STATE> 3
<MEAN> 39
 -4.192688e+000 1.384943e+000 1.529688e+001 -6.152121e+000 2.332955e+000 -1.359070e+000 -2.247794e+000 -3.291312e-001 -2.339166e+000
1.439646e+000 -3.844969e+000 -2.880891e+000 6.652218e+001 1.303583e+000 1.011649e+000 1.645032e+000 1.671035e+000 5.047166e-001 -
1.652937e+000 -1.048727e+000 6.798058e-001 -1.381817e+000 -1.369222e+000 3.608670e-001 2.231764e-002 -5.899567e-001 -2.777144e-001 -
1.489953e-001 -3.124299e-002 -3.336180e-001 -3.611636e-001 2.264947e-001 -4.398102e-001 1.322734e-002 -2.281398e-001 -3.758300e-002
2.433900e-001 2.242147e-001 3.692640e-001
<VARIANCE> 39
 4.205011e+001 6.793561e+000 6.770607e+001 1.000563e+001 3.752867e+001 1.409674e+001 2.057385e+001 1.639126e+001 9.559263e+000
8.444270e+000 8.982311e+000 5.047385e+000 2.873622e+000 7.631572e+000 3.180981e-001 1.297408e+000 1.254034e+000 3.660877e+000
2.944554e+000 5.113333e+000 1.282827e+000 7.595363e-001 1.002803e+000 6.846957e-001 1.396056e+000 1.087084e+000 3.388949e-001
1.878517e-001 2.645418e-001 1.151235e-001 9.468338e-002 1.565774e-001 7.345209e-001 1.862980e-001 3.751449e-002 1.014839e-001
1.483926e-001 1.462476e-001 1.075986e-001
<GCONST> 8.657848e+001
~
```

Poprawienie modeli ciszy (silence models) poprzez dodanie dodatkowego stanu

- krok 7: dodanie dodatkowego przejścia między stanem 2 i 4 oraz 4 i 2 w modelu ciszy oraz dodanie stanu krótkiej pauzy *sp*



Krok 7 - wykonanie

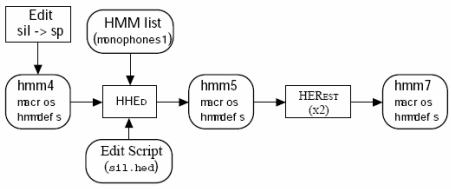
- skopiować środkowy stan modelu ciszy *sil*, aby utworzyć nowy model *sp* i zapisz wynik w pliku hmmdefs czwartej iteracji (katalog hmm4),
- uruchomić moduł edytora HMM - HHEd, aby dodać dodatkowe przejścia między stanami oraz aby podłączyć model *sp* do centralnego stanu modelu *sil*:

```
HHEd –H hmm4/macros –H hmm4/hmmdefs –M hmm5 sil.hed monophones1

sil.hed:
AT 2 4 0.2 {sil.transP}
AT 4 2 0.2 {sil.transP}
AT 1 3 0.3 {sp.transP}
TI silst {sil.state[3], sp.state[2]}
```

- uruchomić dwukrotnie moduł re-estymacji - HERest

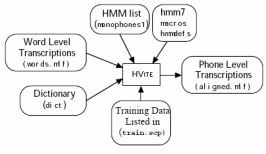
Podsumowanie kroku 7



Uwzględnienie różnej wymowy określonych słów – moduł *HVite*

- krok 8:
  - uruchomić moduł HVite:

```
HVite –l '*' –o SWT –b silence –C config –a –H hmm7/macros \
-H hmm7/hmmdefs –i aligned.mlf –m –t 250.0 –y lab \
-I words.mlf –S train.scp dict monophones1
```



- Ponownie uruchomić dwukrotnie moduł re-estymacji – HERest -> hmm9

Multifony

- krok 9: tworzenie (di)trifonów z monofonów (word internal)

```
HLEd -n triphones1 -l '*' -i wintri.mlf mktri.led aligned.mlf
```

mktri.led:

```
WB sp
WB sil
TC
```

aligned.mlf:

```
#!MLF!#
"*/1.lab"
sil
c6
a
d
sp
z
sp
m
a
\346
e
j
e
m
sp
sil
.
```

wintri.mlf: + **triphones** (wszystkie możliwe)

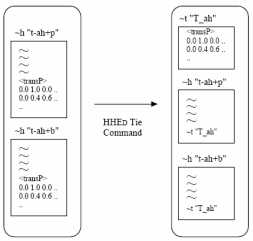
```
#!MLF!#
"*/1.lab"
sil
c6+a
c6-a+d
a-d
sp
z
sp
m+a
m-a+\346
a-\346+e
\346-e+j
e-j+e
j-e+m
e-m
sp
sil
.
```

Tworzenie modeli multifonów

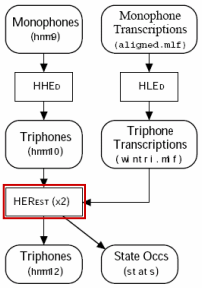
```
HHEd -B -H hmm9/macros -H hmm9/hmmdefs -M hmm10 mktri.hed monophones1
```

mktri.hed:

```
CL triphones1
TI T_Z@ {(*-Z@+*,Z@+*,*-Z@).transP}
TI T_l {(*-l+*,l+*,*-l).transP}
TI T_c6 {(*-c6+*,c6+*,*-c6).transP}
TI T_X {(*-X+*,X+*,*-X).transP}
TI T_Z {(*-Z+*,Z+*,*-Z).transP}
TI T_š {(*-š+*,š+*,*-š).transP}
TI T_a {(*-a+*,a+*,*-a).transP}
TI T_f {(*-f+*,f+*,*-f).transP}
TI T_e {(*-e+*,e+*,*-e).transP}
TI T_d {(*-d+*,d+*,*-d).transP}
TI T_g {(*-g+*,g+*,*-g).transP}
TI T_ć {(*-ć+*,ć+*,*-ć).transP}
TI T_i {(*-i+*,i+*,*-i).transP}
TI T_k {(*-k+*,k+*,*-k).transP}
TI T_j {(*-j+*,j+*,*-j).transP}
TI T_m {(*-m+*,m+*,*-m).transP}
TI T_s6 {(*-s6+*,s6+*,*-s6).transP}
TI T_o {(*-o+*,o+*,*-o).transP}
TI T_n {(*-n+*,n+*,*-n).transP}
...
```



Re-estymacja w kroku 9



Łączenie stanów modeli trifonów

- krok 10:

```
HHEd -B -H hmm12/macros -H hmm12/hmmdefs -M hmm13 tree.hed triphones1
```