

 NoFA-no.md

NoFA -- norsk modell for forced alignment

Innledning

Forced alignment (FA) refererer til algoritmer som tar en lydfil med tale og en ortografisk transkripsjon av talen som input og produserer en fonetisk transkripsjon der hvert segment (hver fon) er tidskodet i lydfila.

FA er nyttig i mange sammenhenger, blant annet fonetikk, fonologi, datalingvistikk og språkteknologisk utviklingsarbeid. For at et FA-verktøy skal fungere for et bestemt språk, trengs en akkustisk modell og et uttaleleksikon for det språket. Språkbanken tilbyr et offentlig tilgjengelig uttaleleksikon for bokmål, [NST-leksikonet](#). Fram til nå har det imidlertid ikke eksistert en akkustisk modell for FA med tilsvarende fri tilgang.

[Nate Young](#) har, på oppdrag av Språkbanken ved Nasjonalbiblioteket, laget en FA-modell for norsk (bokmål), *NoFA*. Denne modellen er laget for FA-verktøyet [Montreal Forced Aligner](#) (MFA). NoFA er trent på Språkbankens taledatabase [NB Tale](#) og den fonetiske delen av RUNDKAST-databasen utarbeida ved NTNU (Amdal et al. 2008). Vi viser til det vedlagte dokumentet *Construction of Forced aligner for Norwegian Bokmål. February Progress Report* for detaljer om utviklinga av denne modellen.

NST-leksikonet er ikke ideelt for FA, for det første fordi det inneholder liten grad av uttalevariasjon og for det meste kun representerer en standarduttale av sør-østnorsk, for det andre fordi det ble utvikla ved årtusenskifte og dermed mangler oppdatert vokabular. I tillegg til en versjon av NST-leksikonet som er lesbart for MFA gjør vi derfor også tilgjengelig et litt utvida leksikon som har blitt brukt i utviklinga av FA-modellen. Vi gjør oppmerksom på at denne er noe eksperimentell og bare i begrensa grad reflekterer uttalevariasjonen i norsk. Begge leksikaene støtter kun bokmål. Språkbanken er i startfasen av et prosjekt der vi skal utvikle uttaleleksika for begge målformer med oppdatert vokabular og dialektal variasjon. Disse vil kunne brukes til FA. For øvrig oppfordres brukere til å gjøre egne oppdateringer og tilpasninger av leksikaene for eget bruk. Om dere ønsker å dele oppdaterte leksika eller oppdateringsregler med Språkbanken, er det kjærkomment. Ta i så fall kontakt med sprakbanken@nb.no. Det er også mulig å automatisk generere uttaleordlister for ord som ikke eksisterer i leksikon ved hjelp av *grafem-til-fonem-konversjon*. Vi kommer tilbake til det under.

Innhold

- **documentation/** inneholder relevant dokumentasjon for NoFA, inkludert dette dokumentet
- **modell/** inneholder den akkustiske modellen for norsk FA, *NoFA_1-0.zip*
- **lexicon/** inneholder leksika formatert for MFA og konvertert til NoFabet (se under). *NST_nob.dict* har alle ortografi-uttale-par i NST-leksikonet. *NoFA-lex.dict* er et leksikon utvikla samtidig med NoFA-modellen.

NoFabet

Et eget sett med ASCII-baserte transkripsjonskonvensjoner, *NoFabet*, er utvikla for dette prosjektet. NoFabet er delvis basert på [ARPABET](#). Hver fon er representert med en streng med én eller flere store bokstaver, tall og andre tegn, og det er mellomrom mellom alle foner.

Lastebilen transkriberes slik i NoFabet: *L AH2 S T AX0 B I13 L NX0*. Alle vokaler og stavelsesbærende konsonanter har et tall som del av transkripsjonen. *1* og *2* står for henholdsvis tonem *1* og *2*. *0* betyr at den stavelsesbærende fonen er trykklett. *3* indikerer bitrykk. Lang vokal transkriberes med to bokstaver, som *I13*, mens kort vokal transkriberes med en *H*, som i *AH2*. For vokaler som *æ*, som ikke fins i ASCII-tegnsettet, er det kun *H* som skiller kort og lang vokal: *AE2* vs. *AEH2*. Tegnet *X* brukes for en del spesielle, stavelsesbærende foner. Stavelsesbærende konsonanter transkriberes med *X*, som *NX0*, og schwa er *AX0*. Under følger en tabell med fonene i NoFabet og X-SAMPA-ekvivalentene i NST-leksikonet:

X-SAMPA	NoFabet	Eksempel
A:	AA0	bad
{:	AE0	vær
{	AEH0	vært
{*l	AEJ0	sei

X-SAMPA	NoFAbet	Eksempel
E*u0	AEW0	sau
A	AH0	hatt
A*I	AJ0	kai
@	AX0	behage
b	B	bil
d	D	dag
e:	EE0	lek
E	EH0	penn
f	F	fin
g	G	gul
h	H	hes
l	IH0	sitt
i:	IIO	vin
j	J	ja
k	K	kost
C	KJ	kino
l	L	land
l=	LX0	
m	M	man
m=	MX0	
n	N	nord
N	NG	eng
n=	NX0	
o:	OA0	rå
O	OAH0	gått
2:	OE0	løk
9	OEH0	høst
9*Y	OEJ0	køye
U	OH0	f*ort
O*Y	OJ0	konvoy
u:	OO0	bod
@U	OU0	show
p	P	pil
r	R	rose
d`	RD	rekord
l`	RL	perle
l`=	RLX0	

X-SAMPA	NoFAbet	Eksempel
n`	RN	barn
n`=	RNX0	
s`	SJ	pers
t`	RT	stort
r=	RX0	
s	S	sil
S	SJ	sju
s=	SX0	
t	T	tid
u0	UH0	russ
u0 j	UH0_J	Anhui
}:	UU0	hus
v	V	vase
w	W	Washington
Y	YH0	nytt
y:	YY0	ny

Grunnen til at NoFAbet er valgt istedenfor X-SAMPA er at fonemtranskripsjonene ligger nærmere opp til norsk ortografi, og det er derfor lettere å lese og manuelt produsere transkripsjoner med denne standarden.

Merk at vi har valgt å gjøre et par forenklinger i konversjonen fra transkripsjonene i det originale NST-leksikonet til NoFAbet-versjonen. For det første har det originale leksikonet markering av stavelsesgrenser, og for flerordsuttrykk, ordgrenser og frasetrykk. Dette har vi valgt å utelate, siden dette ikke er informasjon MFA benytter seg av. For det andre har det originale NST-leksikonet en egen representasjon av retrofleks s, s`, som skiller seg fra den postalveolare frikativene de transkriberer som S. Vi har valgt å bruke SJ for begge fonene.

Hvordan kjøre NoFA

1. Installer MFA. Instruksjoner fins [her](#).
2. Gjør klar dataene. Legg lydfiler og korresponderende ortografiske transkripsjoner i en mappe. Transkripsjonene må være i Praats format *.TextGrid*. Lydfilene må være *.wav*-filer og ha en samplerate på minst 16 kHz. Lydfiler og transkripsjonsfiler må ha samme navn, f.eks. *myrecording.wav* og *myrecording.TextGrid*.
3. Valider filene (kan sløyfes). Stå i mappa der MFA er installert. Kommando: `bin/mfa_validate_dataset sti/til/input sti/til/leksikon sti/til/akkustisk/modell`. Det genereres en tekstfil med ord i transkripsjonene som ikke står i leksikon. I den ferdige FA-transkripsjonen vil disse ordene bli markert som ukjent. Dersom man ønsker at disse skal få en transkripsjon, må man legge dem til i leksikon. Mer om dette under. Valideringsskriptet sier også fra om eventuelle feil i lyd- og transkripsjonsfilene. Se [her](#) for ytterligere detaljer.
4. Kjør aligneren: `bin/mfa_align sti/til/inputfiler sti/til/leksikon sti/til/akkustisk/modell sti/til/output`. FA-transkripsjoner i form av *.TextGrid*-filer vil havne i output-mappa.

For ytterligere detaljer, se [MFAs hjemmeside](#).

Evaluering

Nate Young har testa hvor nøyaktig tidskodning NoFA gir. Testdataene er transkriberte lydfiler fra stortingsforhandlingene med tre talere av tre forskjellige dialekter: Oslo (Anniken Hauglie), Bergen (Audun Lysbakken) og Trøndelag (Trine Skei Grande). Hvert datasett inneholdt mellom 1000 og 1500 foner. Testen viser hvor stort avvik det er mellom tidskodene NoFA produserer og manuelt justerte tidskoder. Testparametre er medianavvik i millisekunder (ms) for fonstart (starttidskode for en fon), prosentdel av predikerte fonstarter som er innafor 20 ms av faktisk fonstart, og prosentdel av predikerte fonstarter som er innafor 10 ms av faktisk fonstart. Resultatene er som følger:

Dialekt	Medianavvik	Prosent innafor 20 ms	Prosent innafor 10 ms
---------	-------------	-----------------------	-----------------------

Dialekt	Medianavvik	Prosent innafor 20 ms	Prosent innafor 10 ms
Oslo	10	79,2%	48,8%
Bergen	10	76,8%	47,8%
Trøndelag	12	72,3%	40,2%

I følge Nate Young er resultatene for Oslo og Bergen innafor gullstandarden for FA-systemer, og for Trøndelag nær gullstandarden. Se den vedlagte rapporten *Construction of Forced aligner for Norwegian Bokmål, Final Report* for detaljer.

Begrensninger ved leksikon

Leksika som er kompatible med MFA er tekstfiler der hver linje har en ortografisk form av et ord og en transkripsjon. Den ortografiske formen og transkripsjonen er skilt med mellomrom. Dersom et ord har flere alternative transkripsjoner, vil disse være egne linjer med samme ortografiske form, men ulik transkripsjon. MFA vil kun produsere transkripsjoner som står i leksikon. Dersom et ord ikke står i leksikon, vil MFA sette inn en *ukjent*-tagg. Dersom ordet står i leksikon, men med en annen uttale enn den som faktisk forekommer i lydfila, vil transkripsjonen ikke fullt ut reflektere det som sies.

NST-leksikonet er laga rundt årtusenskiftet, og det vil derfor være en del nyere vokabular som ikke står i leksikon. Dessuten tillater norsk at man setter sammen substantiver til nye, sammensatte substantiver. De aller fleste transkripsjoner vil derfor trolig inneholde ord som ikke står i leksikon. Dersom man ønsker at disse skal tas hensyn til av MFA, må man legge dem til i leksikon. Man kan enten skrive inn orda og transkripsjonene manuelt, eller man kan bruke såkalt *grafem-til-fonem-konversjon* (GtF). Språkbanken har trent en modell for GtF-systemet [Phonetisaurus](#) ved hjelp av NST-leksikonet. Den kan man finne [her](#).

NST-leksikonet er basert på østlandsk, og MFA vil derfor produsere bedre transkripsjoner for østlandsk enn andre dialekter. NST-leksikonet inneholder heller ikke kontraksjoner o.l., som man vil finne i rask tale. I *NoFA-lex.dict* har vi i noen grad forsøkt å bøte på dette. Brukere som trenger nøyaktige transkripsjoner oppfordres likevel til å oppdatere leksikon selv. Språkbanken jobber med å lage uttaleleksikon med dialektal variasjon for bokmål og nynorsk, som vil kunne brukes med MFA.

Bruksvilkår og kontaktinformasjon

NoFA-modellen og tilhørende ressurser er offentlig eiendom med lisensen [CC0](#). Det er med andre ord ingen bedrensninger på bruken av dem. Se [her](#) for MFAs lisens.

Denne ressursene deles som de er, og kommer ikke med noen garanti av noe slag. Vi er imidlertid interessert i å få vite om det dersom brukere finner feil, har forslag til forbedringer eller har spørsmål. Vi kan nås på sprakbanken@nb.no.

Referanser

- I. Amdal, O.M. Strand, J. Alberg, T. Svendsen. 2008. RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus. Proceedings of LREC 2008