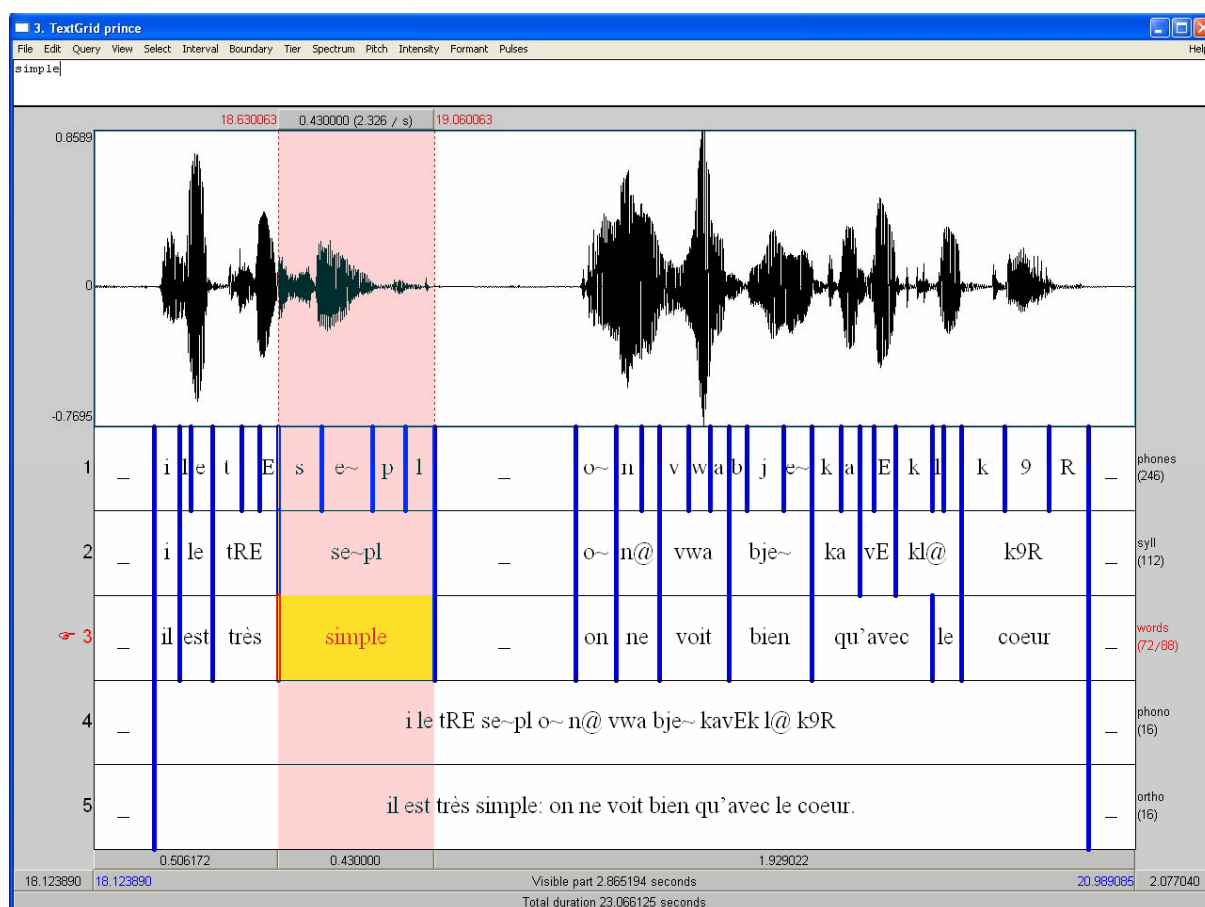


Tutoriel EasyAlign

Jean-Philippe Goldman- avril 2010

Ce document décrit les étapes successives pour l'installation et l'utilisation de EasyAlign.

EasyAlign est un plugin (extension) de Praat qui permet de créer de manière semi-automatique une annotation multi-tires contenant une segmentation alignée en phonèmes, en syllabes, en mots et en énoncés, à partir d'un enregistrement de parole et d'une transcription orthographique de l'enregistrement. La procédure se déroule en une succession de 3 étapes automatiques entre lesquelles des réajustements manuels peuvent être nécessaires. Le résultat est un TextGrid de 5 tires *phones*, *syll*, *words*, *phono* et *ortho* comme illustré ci-dessous. EasyAlign est conçu pour le français et ne fonctionne que dans un environnement Windows.

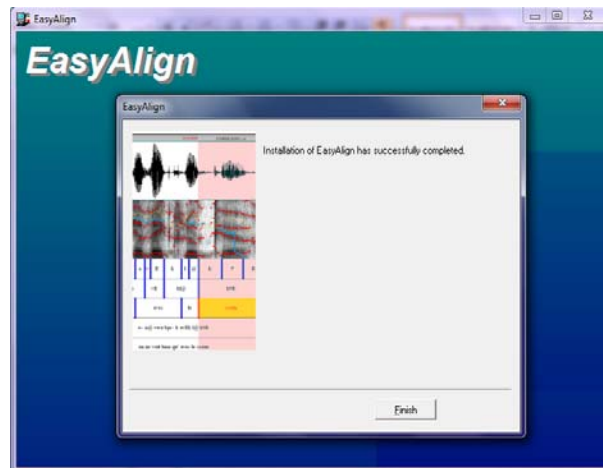


I. Installation et démonstration.....	2
Installation.....	2
Déroulement de la démonstration.....	3
1. Macro-segmentation	3
2. Phonétisation	4
3. Segmentation en phonèmes	5
II. Prise en main détaillée d’EasyAlign.....	5
1. Procédure usuelle : segmenter un son transcrit dans un fichier externe	5
2 Pré-formater une transcription en « une phrase par ligne »	9
Segmenter un enregistrement non transcrit	9
Outils supplémentaires	9
III. Questions fréquentes	9
Comment EasyAlign effectue la macro-segmentation ?	9
Comment EasyAlign effectue la phonétisation ?	9
Comment EasyAlign effectue la segmentation ?.....	10
Comment EasyAlign effectue la syllabation ?	10
Qu’est-ce que l’alphabet SAMPA ?.....	10
Où se trouvent les scripts de EasyAlign ?.....	10
Pourquoi les frontières de syllabes ne correspondent-elles pas à celles des mots ?	11
Out of memory lors de l’étape 3	11
V. Liens	12

I. Installation et démonstration

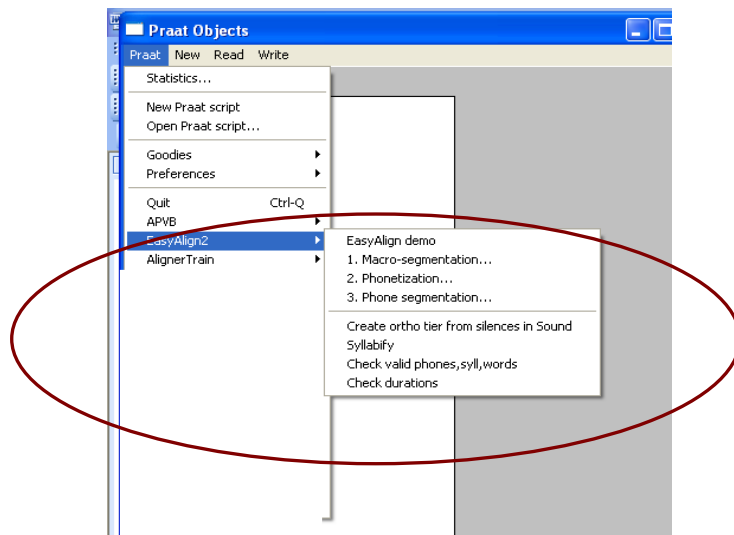
Installation

Télécharger et exécuter le fichier EasyAlignSetup.exe depuis l’adresse latlcui.unige.ch/phonetique .
 Durant l’installation, appuyer deux fois sur Next puis Finish pour terminer.



S'assurer d'avoir une version récente de Praat et le démarrer (ou le redémarrer s'il était actif pendant l'installation d'EasyAlign), un sous-menu EasyAlign apparaît dans le menu Praat.

Une démonstration est incluse pour présenter les étapes une à une.



Déroulement de la démonstration

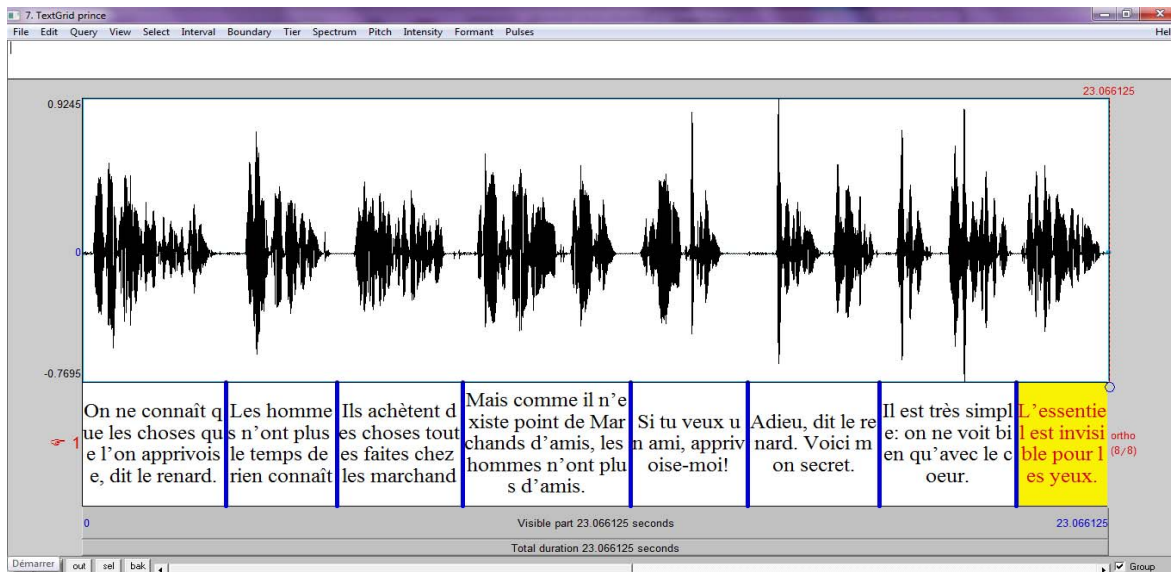
Lancer la démonstration avec Praat → EasyAlign → EasyAlign Demo

Un objet « Sound prince » et un objet « Strings prince » se chargent. Il s'agit d'un son provenant d'un fichier sonore **prince.wav** qui est une lecture orale par une locutrice, et du contenu d'un fichier texte **prince.txt** correspondant à la transcription avec le format d' « une phrase par ligne ».

Appuyer sur Continue après chaque étape.

1. Macro-segmentation

La première étape automatique de EasyAlign crée une macro-segmentation sous la forme d'un TextGrid contenant une tire nommée *ortho*. Elle nécessite en entrée le son à segmenter (qui est chargé comme un objet Sound) et un fichier texte qui contient la transcription orthographique sous forme d'énoncés (qui est chargé comme un objet Strings).

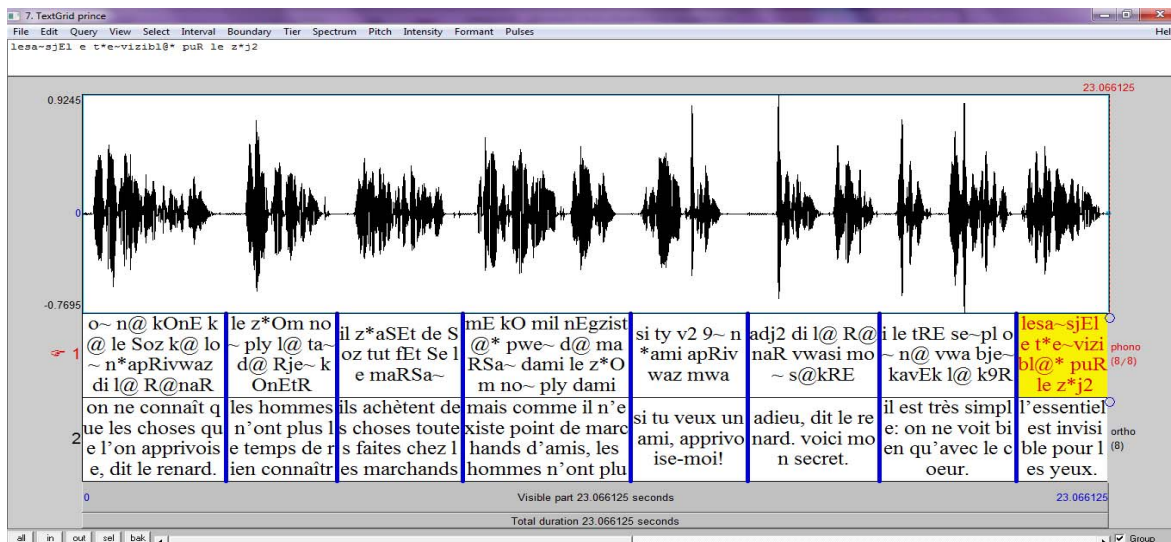


L'utilisateur doit alors **vérifier par l'écoute la position de chacune des frontières**, et éventuellement ajuster manuellement, afin que chaque intervalle corresponde à la transcription. En général, il est pratique de régler la fenêtre d'affichage du signal sur une largeur de 10 secondes environ, c'est-à-dire plus serré que sur l'illustration ci-dessus.

Pour cette démonstration, les frontières sont placées correctement. Il n'y a aucun ajustement à faire.

2. Phonétisation

La seconde étape automatique crée une tire *phono*, c'est-à-dire une tire avec les mêmes frontières, et y remplace la transcription orthographique en une transcription phonétique selon l'alphabet phonétique SAMPA (cf. Questions fréquentes). Voici le TextGrid augmenté de la tire *phono*.



L'utilisateur doit alors **vérifier par l'écoute la transcription phonétique** de chaque intervalle. Une manière de faire consiste à vérifier par l'écoute la transcription de chaque intervalle. Il s'agit notamment de prêter attention aux variations phonologiques comme les liaisons, les élisions, la prononciation des nombres et des sigles (lecture ou épellation, e.g. O.N.U. [ony / oEny]). Une étoile est insérée aux liaisons et à certains schwas.

Dans cet exemple, l'unique différence entre la transcription phonétique et la prononciation de la locutrice se trouve à l'énoncé 3 avec l'épenthèse d'un schwa à la fin du mot « toutes ». Il convient de rajouter manuellement ce schwa (dont le symbole est [ə]), en prenant garde à ne pas rajouter ni enlever des espaces (cf Segmenter un fichier étape.5, et erreurs infra).

3. Segmentation en phonèmes

La troisième étape a pour but la création des tires *phones*, *syll* et *words*. Un moteur de reconnaissance de la parole est exécuté pour repérer simultanément les frontières temporelles de chaque mot et de chaque phonème. Dans un deuxième temps, les frontières des syllabes sont calculées au moyen de règles phonologiques basées sur la sonorité.

La première figure de ce document illustre le résultat final. L'utilisateur peut écouter chaque mot, chaque syllabe ou chaque phonème pour vérifier et éventuellement ajuster les frontières temporelles.

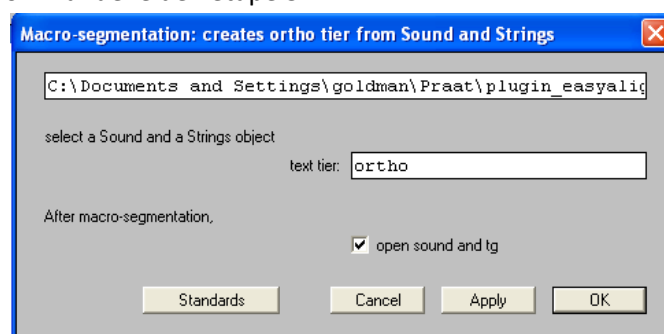
Dans la démonstration, tous les énoncés ont pu être alignés correctement.

II. Prise en main détaillée d'EasyAlign

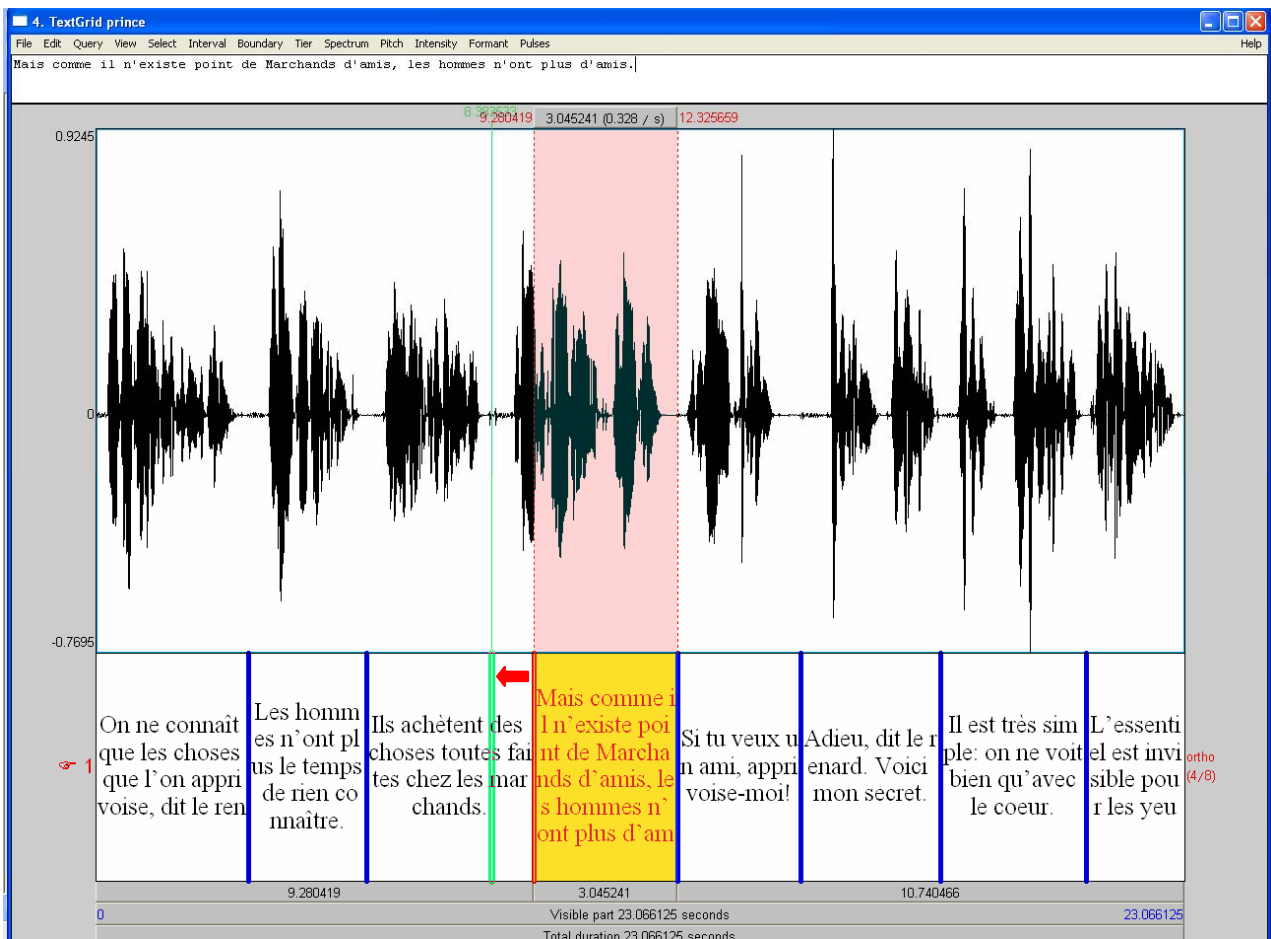
Le déroulement de la procédure complète pour segmenter un enregistrement sonore dépend de la transcription associée. Cette dernière peut être contenue dans un fichier annexe. Si elle est sous le format commun d'une « phrase par ligne », alors il s'agira alors de suivre la procédure usuelle décrite ci-dessous (II.1). Si la transcription est en paragraphes, il faudra d'abord procéder à un pré-formatage du texte (II.2), puis suivre la procédure usuelle en (II.1). Finalement, si la transcription est absente, l'expérience montre qu'il est plus efficace de transcrire directement dans Praat en créant la tire *ortho* manuellement comme décrit en (II.3) puis de commencer la segmentation directement à l'étape 4 de la procédure usuelle (II.1).

1. Procédure usuelle : segmenter un son transcrit dans un fichier externe

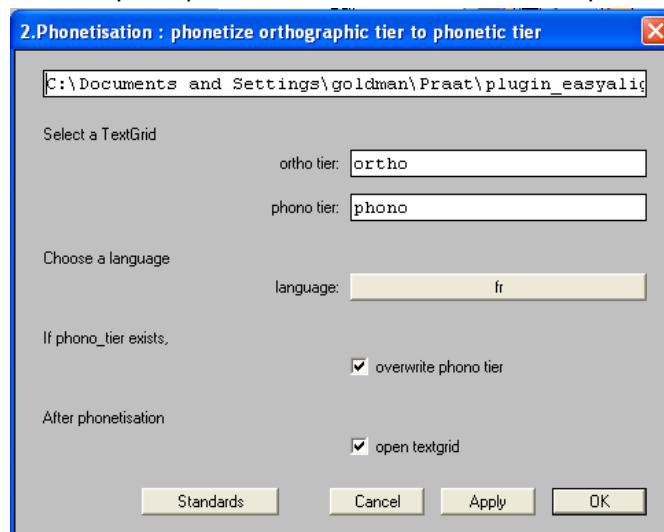
1. **Ouvrir un fichier sonore** (objet Sound) **et une transcription orthographique** pré formatée comme « une phrase par ligne » (cf ci-dessous). Attention à bien ouvrir le son selon un objet Sound et non LongSound.
2. Sélectionner ces deux objets (Sound et Strings) et **exécuter le script 1. Macro-segmentation** (menu Praat → menu EasyAlign). Un TextGrid est créé et contient une tire nommée *ortho* par défaut. Après l'exécution du script, le son et le TextGrid créé sont ouverts conjointement pour la vérification manuelle de l'étape 3.



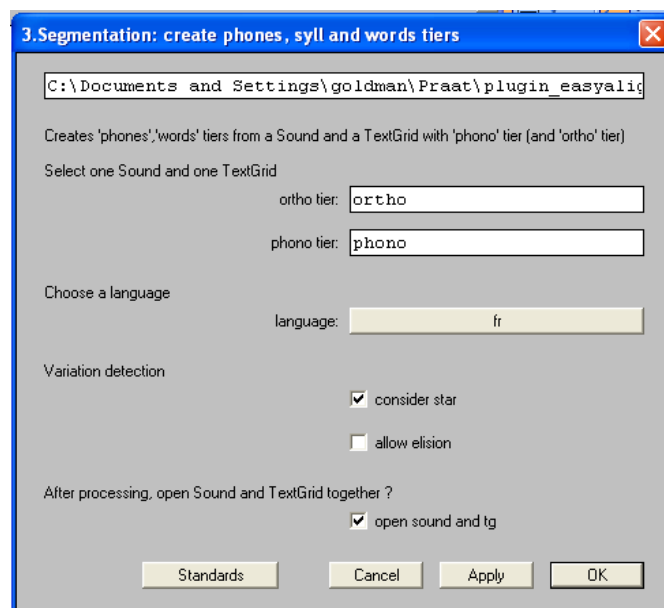
3. Dans le TextGrid créé, **vérifier par l'écoute la position de chaque frontière** temporelle de la tire *ortho* et le cas échéant, il faut les ajuster. Dans l'image ci-dessous, la frontière rouge a été mal positionnée par l'étape 2, il faut la déplacer vers la gauche (position verte) pour s'ajuster à la pause entre les deux phrases.



4. Sélectionner le TextGrid et **exécuter le script 2. Phonétisation**. Par défaut, la tire nommée *ortho* sera phonétisée en *phono* par un système automatique de conversion graphème-phonème, basé sur la transcription orthographique uniquement. Vérifier le nom des 2 tires, la langue choisie. Une option permet d'écraser ou non une tire phono existante.



5. Dans le TextGrid, **vérifier et corriger dans la tire *phono* la transcription phonétique.** Parmi les variations possibles, on peut citer :
- Les erreurs de transcription (on peut alors aussi corriger la transcription)
 - les erreurs de prononciation
 - les liaisons
 - les élisions de schwa ([sam(ə)di]) mais aussi d'autres phonèmes ([i(l)zɔ̃di])
 - les épenthèses de schwa (toutes faites [tut@fɛt])
 - la prononciation des nombres (dix-neuf cent cinquante vs. mille neuf cent cinquante, ou septante vs. soixante-dix)
 - la prononciation des sigles (lecture vs. épellation, O.N.U. [ony / oEny]).
 - Une étoile est insérée aux liaisons et à certains schwas.
- ➔ Attention, il est impératif de respecter les caractères d'espace pour maintenir le même nombre de « mots » dans les deux tires *ortho* et *phono*. Par exemple « qu'avec le cœur » transcrit comme [kavɛk l@ k9R] ou « qu' avec le cœur » transcrit comme [k avɛk l@ k9R].
6. Sélectionner le son et le TextGrid et **exécuter le script 3. Segmentation.** Vérifier le nom des 2 tires, la langue choisie. Les deux cases à cocher suivantes indiquent au système d'alignement de prendre en charge la décision du maintien ou non de certains phonèmes (ceux marqués d'une étoile pour la case « consider star », et les schwa pour la case « allow elision »).



7. **Vérifier** dans la fenêtre Info de Praat **que tous les énoncés ont bien été alignés.** Si ce n'est pas le cas, corriger le TextGrid selon les cas ci-dessous puis ré exécuter le script 3. Segmentation. Plusieurs cas peuvent se présenter:
- **Caractères non SAMPA dans la transcription phonétique** : la transcription de la tire *phono* contient des caractères qui n'appartiennent pas à l'alphabet SAMPA. Pour cela il convient

de modifier la tire *phono*. Attention, si le script 2.Phonetisation est exécuté à nouveau, les modifications manuelles dans la tire *phono* seront écrasées.

- Le nombre de mots différent pour les tires *phono* et *ortho* : il se peut que durant la vérification manuelle de la transcription phonétique, un caractère d'espace ait été supprimé ou ajouté. Il convient alors de rétablir cette correspondance de nombre de mots sur les tires *ortho* et *phono*. Puis de relancer le script 3.Segmentation. Si la différence de nombre de mots ne provient pas d'une correction manuelle mais de l'étape de phonétisation automatique, il convient de la corriger quand même et de le signaler à l'auteur (jeanphilippegoldman@gmail.com)
- La segmentation n'a pu être faite car l'énoncé est trop long, ou la transcription phonétique est trop éloignée de la prononciation, ou le signal sonore est de mauvaise qualité ou encore le débit de parole est trop rapide. Plusieurs pistes de solutions :
 - vérifier les transcriptions orthographiques et phonétiques de l'énoncé non aligné et les corriger le cas échéant. Attention à ne pas relancer, à ce stade, le script 2.Phonétisation, sous peine d'écraser des modifications manuelles de la transcription phonétique antérieures.
 - scinder l'énoncé en plusieurs énoncés en insérant une frontière dans les tires ortho ET phono. Il faut s'assurer que les frontières coïncident exactement (Dans TextGrid Editor, utiliser le menu Boundary→Add on selected tier). Puis relancer le script 3.Segmentation

Il se peut que malgré ces efforts, EasyAlign ne puisse proposer un alignement en phonèmes. Si l'alignement est indispensable, il pourra être fait manuellement pour les tires *phones* et *words*. La syllabation pourra se faire automatiquement avec l'outil Syllabify du menu EasyAlign.

8. **Vérifier manuellement et ajuster la position des frontières temporelles** des phonèmes, syllabes et mots. En cas de déplacement de frontières, il convient de respecter l'intégrité de frontières simultanées en les déplaçant ensemble. Pour ce faire, il faut déplacer les frontières en appuyant sur la touche Shift en même temps. Il est possible de changer ce comportement de Praat, en modifiant dans le menu Praat → Preferences, l'option « With the shift key, you drag multiple boundaries/ a single boundary »

9. Ne pas oublier de **sauver le TextGrid**

Attention certains ajustements manuels seront perdus si une étape automatique est ré-exécutée par la suite, notamment :

- si le script 2.Phonétisation (étape4) est ré-exécuté après des modifications manuelles de transcription phonétique dans la tire phono (étape 5), celles-ci seront écrasées
- si le script 3.Segmentation (étape7) est ré-exécuté après des ajustements manuels de frontières de phonèmes (étape8), ceux-ci seront perdus.

Il faut donc procéder de manière progressive dans l'enchaînement des étapes décrites ci-dessus.

2 Pré-formater une transcription en « une phrase par ligne »

Pour faire en sorte que chaque ligne contienne la transcription d'une seule phrase, il faut insérer un « retour à la ligne » après chaque phrase, ou mieux, entre chaque énoncé terminé par une pause. Ceci peut se faire dans n'importe quel éditeur de texte. Il faut éviter si possible, les énoncés contenant plus de 50 mots. Ainsi dans l'exemple du Petit Prince de la démonstration incluse, le fichier texte se présente ainsi :

```
On ne connaît que les choses que l'on apprivoise, dit le renard.  
Les hommes n'ont plus le temps de rien connaître.  
Ils achètent des choses toutes faites chez les marchands.  
Mais comme il n'existe point de Marchands d'amis, les hommes n'ont plus d'amis.  
Si tu veux un ami, apprivoise-moi!  
Adieu, dit le renard. Voici mon secret.  
Il est très simple: on ne voit bien qu'avec le coeur.  
L'essentiel est invisible pour les yeux.
```

Il faut veiller également à supprimer des caractères ou des portions de texte qui ne correspondent pas à ce qui est prononcé, comme par exemple des balises à chevrons (<html>) ou autres signes muets. La ponctuation usuelle peut néanmoins être gardée.

Dans l'éditeur de texte, ne pas oublier d'enregistrer ces modifications dans un le fichier de format texte brut (généralement une extension .txt et non .doc ou .rtf).

Segmenter un enregistrement non transcrit

Pour cela, il faut ouvrir le fichier sonore dans Praat (menu Read → Read from file...), puis créer un nouveau TextGrid (Annotate → To TextGrid... avec les options All tier names : ortho et Point tiers : laisser vide). Le TextGrid créé sera de la même durée que le fichier sonore et aura une tire ortho vide.

Puis, il faut ouvrir le son et ce TextGrid, apposer des frontières de préférences vers le milieu des pauses et transcrire chaque intervalle un par un en les écoutant autant de fois que nécessaire. Ne pas dépasser 5 secondes comme longueur d'intervalle pour éviter une transcription fastidieuse.

Ne pas oublier de sauvegarder le TextGrid (Menu Praat → Write → Write to text file...) puis effectuer la procédure de segmentation à partir de l'étape de phonétisation en exécutant le script 2.Phonetisation.

Outils supplémentaires

A décrire...

III. Questions fréquentes

Comment EasyAlign effectue la macro-segmentation ?

L'algorithme interne se base une mise en relation de la durée du son à aligner et de la longueur de la transcription pour estimer la position des frontières d'énoncés, tout en privilégiant les pauses.

Comment EasyAlign effectue la phonétisation ?

Un composant issu du système de synthèse à partir de texte eLite développé à la Faculté Polytechnique de Mons, effectue une analyse linguistique de la transcription orthographique pour produire une transcription phonétique à partir d'un dictionnaire et de règles de prononciation.

Comment EasyAlign effectue la segmentation ?

Pour chaque énoncé, les transcriptions orthographique et phonétique permettent de créer un mini-dictionnaire de prononciation qui est utilisé par un moteur de reconnaissance de la parole HTK (HMM Tool Kit) réglé en mode « alignement forcé » pour repérer les frontières temporelles des phonèmes et des mots.

Comment EasyAlign effectue la syllabation ?

Les deux principes suivis sont « une et une seule voyelle par syllabe » et le principe de sonorité pour le découpage des groupes de consonnes. Les pauses sont également utilisées comme frontières de syllabes.

Qu'est-ce que l'alphabet SAMPA ?

SAMPA est un alphabet phonétique qui diffère peu de l'Alphabet Phonétique International (API) et qui présente l'avantage de n'utiliser que des caractères simples. Voici cet alphabet avec la correspondance API et des exemples. Les différences entre SAMPA et API sont grisées.

SAMPA	IPA	Exemple	SAMPA	IPA	Exemple
P	p	Pas	i	i	Riz
T	t	Tas	ʏ	y	Rue
k	k	Cas	u	u	Eoue
b	b	Bas	o	o	Rôt
d	d	Dada	O	ɔ	Robe
g	g	Gars	ʌ	a/a	Ras
f	F	Fa	e	e	Rez
s	S	Sa	ɛ	ɛ	Raie
S	ʃ	Chat	2	ø	Deux
V	v	Va	9	œ	Neuf
Z	z	Zut	@	ə	Cheval
Z	ʒ	Gite	o~	õ	Bon
L	l	Las	a~	ã	Banc
R	ʀ	Ras	e~	ẽ / œ̃	Bain
M	m	Ma			
N	n	Na !			
N	ŋ	Tang			
J	j	Hier			
W	w	Ouate			
H	ɥ	Huit			

Où se trouvent les scripts de EasyAlign ?

L'installateur crée un répertoire plugin_easyalign dans le répertoire que Praat crée lui-même. En général, ce répertoire est **C:\Documents and Settings\\Praat\plugin_easyalign**

Pourquoi les frontières de syllabes ne correspondent-elles pas à celles des mots ?

La tier *syll* est calculée dans un second temps, après les tiers *phones* et *words*. La syllabation est réalisée en regroupant les phonèmes selon le principe de sonorité et en considérant une et une seule voyelle par syllabe. Les enchaînements et élisions peuvent donner lieu à une segmentation syllabique surprenante. Par exemple :

des choses issues de radios commerciales ou de radios classiques *et caetera...*

[dɛ-ʃo-zi-sy-dʁa-dʒo-kɔ-mɛr-sja-lu-də-ʁa-dʒo-klɑ-si-kɛk-sɛ-tɛ-ʁa]]

Out of memory lors de l'étape 3

L'étape 3 commence par ré échantillonner le son à 16000 Hz pour qu'il soit dans le même format que les modèles acoustiques utilisés pour l'alignement. Pour cela, Praat utilise un algorithme très précis mais coûteux en puissance de calcul. Cette erreur surgit si le fichier sonore fait plus de dix minutes. Ce seuil varie en fonction de la puissance de l'ordinateur utilisé.

La solution consiste à rééchantillonner (*resample*) le fichier sonore en 16000Hz dans un autre logiciel de traitement sonore comme CoolEdit ou Goldwave.

IV. Liens

Praat	Logiciel libre d'analyse de la parole (Paul Boersma & Daniel Weenink) www.praat.org
Autour de Praat	http://latlcui.unige.ch/phonetique (prise en main de Praat pdf 8 pages) http://pierre.free.fr/indexpraat.htm (mesure acoustique, web en 9 pages) http://www.uclouvain.be/cps/ucl/doc/valibel/documents/utiliserPraat.pdf (prise en main ; pdf de 12 pages) http://www.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf (pdf de 27 pages) http://person2.sol.lu.se/SidneyWood/prate/frames.html (web) http://uk.groups.yahoo.com/group/praat-users/ (groupe de discussion Yahoo)
Transcription	SAMPA Alphabet phonétique http://www.phon.ucl.ac.uk/home/sampa/french.htm IPA Alphabet phonétique international http://www.arts.gla.ac.uk/IPA/ Conseil pour l'encodage de caractères phonétiques http://latlcui.unige.ch/phonetique Convention de transcription VALIBEL et tutoriel http://www.uclouvain.be/cps/ucl/doc/valibel/documents/transcrirePraat.pdf
Prosogram	Extension de Praat pour la transcription prosodique, utilisant une stylisation du contour mélodique fondée sur un modèle de perception tonale et sur une segmentation automatique en noyaux syllabiques (Piet Mertens) http://bach.arts.kuleuven.be/pmertens/prosogram/
C-Prom	Corpus libre de parole transcrite, annotée et segmentée http://sites.google.com/site/corpusprom
Analor	Logiciel libre d'étiquetage et de séquençage intonatifs, basés sur l'analyse prosodique du discours (Bernard Victorri - Mathieu Avanzi - Anne Lacheret-Dujour) http://www.lattice.cnrs.fr/-Analor-