

Annotation Manual – Version 1.3

This manual was last updated on 2016-05-02

CorpAGEst (2013-2015): “A corpus-based multimodal approach to the pragmatic competence of the elderly”

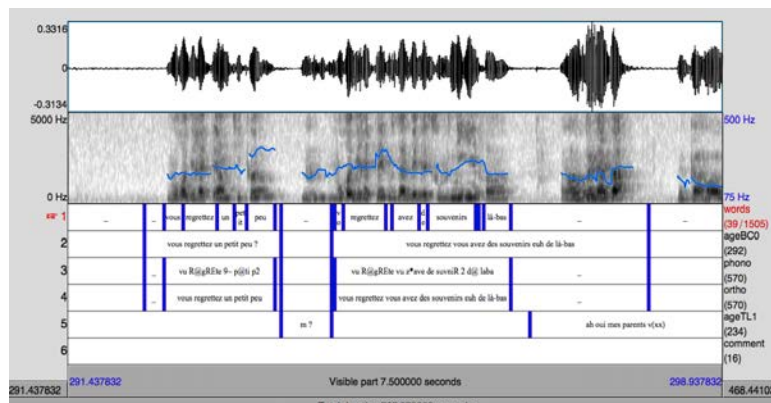
People Marie Curie Actions (PIEF-GA-2012-328282)



Multimodal annotation guidelines

II. Speech annotation guidelines

(Praat program / EasyAlign plugin / ELAN software)



Original Author: Catherine Bolly
With the collaboration of: Julie Kairet

Contact: catherine.bolly@uclouvain.be; catherinebolly@hotmail.com

Web: <http://corpagest.org>

Logo: © The Shelf Company <http://www.theshelf.fr>

Content

Introduction	2
☞ The CorpAGEst project.....	2
☞ Multimodal annotation: from form-based to function-based analysis.....	2
II. Speech annotation guidelines	4
General principles.....	4
1. Speech transcription (<i>Praat</i> software)	4
Transcription principles	4
1.1. Tier structure in <i>Praat</i>	5
1.2. Conventions of transcription	5
Transcription principles	5
1.3. Text/sound anonymization	9
Anonymization principles.....	9
1.4. Transcription revision	10
1.4.1. Revision <i>with</i> the sound signal	10
1.4.2. Revision <i>without</i> the sound signal	10
1.5. Transcription export (<i>Transformer</i> program)	10
1.5.1. Visualization of the raw text output.....	11
1.5.2. Visualization of the XML output.....	11
2. Text-sound alignment	12
2.1. Automatic alignment (<i>EasyAlign</i> plugin)	12
2.1.1. Creation of a “simplified” textgrid.....	12
2.1.2. Isolation of each speaker in one textgrid	13
2.1.3. How to use the EasyAlign plugin	13
2.2. Manual verification	15
2.2.1. Toward the final textgrid	15

Introduction

☞ The CorpAGEst project

The CorpAGEst project (“A corpus-based multimodal approach to the pragmatic competence of the elderly”) aims at establishing the gestural and verbal profile of very old people in aging, looking at their pragmatic competence from a naturalistic perspective. The CorpAGEst assumption is that multimodal (inter)subjective markers of stance are highly relevant cues for the measurement of communicative competence in later life. The project aims *in fine* at a better understanding of the way in which the verbal and gestural dimensions interact to make sense in real-world settings (thus going far beyond the specific scope of the present project). This project has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° [PIEF-GA-2012-328282].

The CorpAGEst corpus (Bolly & Boutet, forthcoming) is comprised of face-to-face conversations between an adult and a very old subject (75 y. old and more) living at home or in a residential home. The corpus data consist of semi-directed interviews, which have been audio and video recorded, transcribed and aligned to the sound signal. The corpus is two-fold, including transversal and longitudinal subcorpora. Contextual independent variables are part of the corpus design, namely environment (private vs. residential home), the social tie between the participants (familiar vs. unknown interviewer) and the task type (focusing on past events vs. present-day life). The corpus is part of the international [CLARe initiative](#) (“Corpora for Language and Aging Research”), which combines methods in linguistics and issues in aging, and advocates for more corpus-based “naturalistic” approaches in the field.

The multimodal data (text, sound and video) were aligned to the sound signal in partition mode, using the Praat program (Boersma & Weernink, 2014), the EasyAlign plugin (Goldman, 2011), and the ELAN software (Wittenburg *et al.*, 2006). The transcription standards adopted for the oral component were slightly adapted from those of the [Valibel research center](#) (Dister *et al.*, 2007 [2009]), as described in the part of the manual dedicated to speech.

For further detail, see the project website: <http://corpigest.org>.

☞ Multimodal annotation: from form-based to function-based analysis

The perspective adopted in the CorpAGEst project is a form-based one (see Müller *et al.*, 2013), extended and applied to facial expressions, gaze, hand gestures, and body gestures (*viz.* head, shoulders, torso, legs, feet). Notably, the annotation procedure lays on a triple principle, according to which a visible action is considered as a *potentially* meaningful gesture unit in the ongoing flow of interaction: (i) the “visibility” criterion: identification of all actions that are visible in the interaction flow, through the eyes of the camera recorder and through those of the analyst; (ii) the “meaning potential” criterion: from the semantic-pragmatic perspective, every visible action identified must potentially convey one semantic-pragmatic meaning in the particular context of its realization (thus also including beats, adaptors, deictic, and interactive gestures), from the point of view of the analyst; (iii) the “formal distinctiveness” criterion: to distinguish between consecutive moves in the interaction flow, there must be at least one change in formal/physiological parameters (e.g. shape for the hand, direction of the head, etc.), by comparison with the preceding and following gesture phase or move.

At the level of speech, the protocol for discourse markers identification and annotation follows the one developed within the [MDMA project](#) (“Model for Discourse Marker Annotation” – see Bolly *et al.*, 2015; Bolly *et al.*, forthc.). The methodology of MDMA starts from an independent selection of candidate discourse markers by several expert coders, which then undergo syntactic and semantic description through an operational annotation model. A specific section of the manual (still in progress) is dedicated to speech transcription (via Praat), alignment (via EasyAlign), and annotation (in ELAN).

Starting with mono-modal analyses (gesture vs. speech) and focusing on one group of articulators at a time within each modality (*viz.* face, gaze, head, shoulders, torso, hands, legs, and feet – for the nonverbal mode), the annotation procedure next moves to a multimodal and functional perspective on pragmatic cues (*viz.* emotions and (non)verbal pragmatic markers. The model for the annotation of pragmatic functions is part of the MDMA project (see above) and is a collaborative work (see Bolly & Crible, Antwerp 2015), which has been developed to be transferrable to several modalities and languages (see Bolly *et al.*, Göttingen 2015).

Modality: nonverbal/gesture	Articulators
1. Facial displays	
	Eyebrows
	Eyes
	Gaze
	Mouth
2. Hand gestures	
	Hands
3. Body gestures	
	Head
	Shoulders
	Torso
	Legs
	Feet
Modality: verbal/speech	Levels of analysis
	Pragmatic markers
FUNCTION-BASED ANALYSIS - Multimodal annotation of emotions - Multimodal annotation of pragmatic functions	

Table 1. Form-based and function-based approach to corpus data in CorpAGEst

II. Speech annotation guidelines

General principles

How to name speech files, step by step, during data treatment in order to keep traces of the work in progress? (This is a suggestion, not a recommendation, to adapt according to research needs...)

ageBN1r-1_S2_transcript_Ju_20150313_EC	Ongoing phase of transcription (“EC”= “En Cours”) by Julie (“Ju”). File lastly saved on 13/03/2015. Recording, working data: second sample (“S2”) of the first interview (“r-1”) with Nadine (“ageBN1”).
ageBN1r-1_S2_transcript_ok	Transcription of the sample done (“ok”) and <i>ideally</i> revised by a second coder.
ageBN1r-1_S2_align_Ju_20150313_EC	Ongoing phase of alignment (“EC”= “En Cours”) by Julie (“Ju”).
ageBN1r-1_S2_align_ok	Automatic alignment finished, including the <i>a posteriori</i> manual track of auto-generated errors (e.g. “mm” erroneously transformed into “millimeter” by the EasyAlign plugin).
ageBN1r-1_S2_L1L2_aligned_ok	Output file resulting from the alignment procedure.
ageBN1r-1_S2_L1_aligned_ok	Output file created to obtain one file per speaker (among others, useful for in-depth prosodic analyses). Here, the L1 speaker refers to the older speaker (viz. the informant).
ageBN1r-1_S2_L2_aligned_ok	Output file created to obtain one file per speaker (among others, useful for in-depth prosodic analyses). Here, the L2 speaker refers to the younger speaker (viz. the interview conductor).

1. Speech transcription (*Praat* software)

The sound samples will be transcribed using the software *Praat*. The use of high-quality headphone is required.

Transcription principles

How to create a file in Praat, that is, how to create a new “Textgrid”?

- 1) Open the file in *Praat* (“Open” > “Read from file”)
- 2) Create a Textgrid (“Annotate” > “Textgrid”)
- 3) Open the Textgrid and the sound by selecting together the textgrid and the sound (> “View & Edit”)
- 4) Create 5 tiers (“Tier” > “Add interval tier”, see section 1.1)

How to create an “interval”, that is, how to create an annotation span?

In the targeted tier, click to create a boundary and move it to create an interval. The boundaries are put to create an interval corresponding *grasso modo* to one meaning unit at the level of utterances. Note that there is no theoretical implication here, as this is a technical preliminary step in the data transcription process.

How many speakers at a time?

It is recommended to transcribe the speech of one speaker at a time. But this is not a rigid rule, as it is sometimes more efficient to take into account the two speakers at the same time, especially in the case of overlaps.

How to save my work?

!!!Be careful: Praat does not automatically save your work!!! Thus frequently save your work.

How long must be a transcription interval?

Usually, it is recommended not to exceed 10 seconds per interval to be transcribed (but this is an approximation, of course).

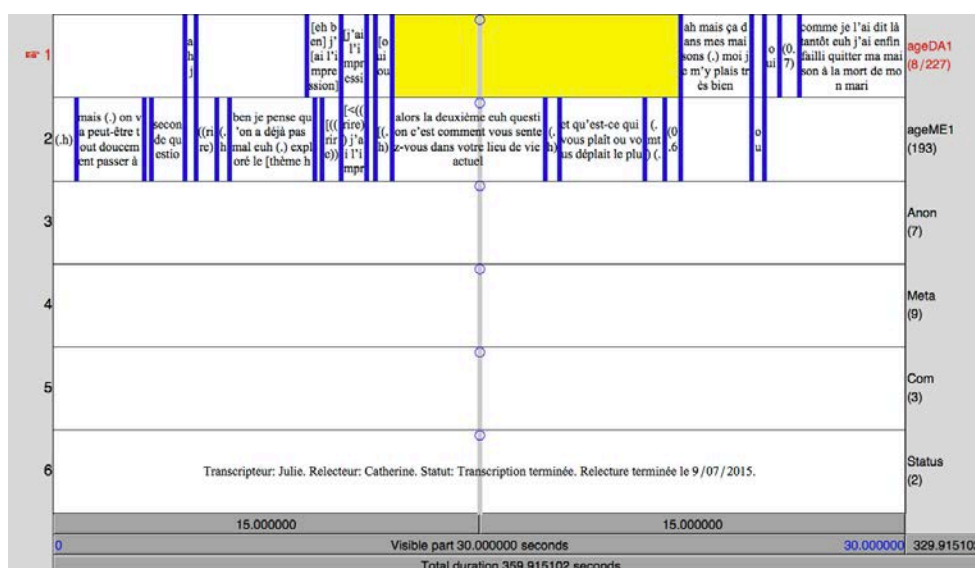
How can I use the Praat functionalities (boundaries, shortcuts, principles, etc.)?

See useful existing papers (Goldman, etc.).

1.1. Tier structure in Praat

Tiers are structured as follows in every TexGrid:

First speaker code [agexxx]	Orthographic transcription of the first speaker speech (usually the interviewee)
Second speaker code [agexxx]	Orthographic transcription of the second speaker speech (usually the interviewer)
Segment to be anonymized [Anon]	Delimitation of the segment that will be anonymized (recognizable thanks to the “#” symbol preceding the targeted segment)
Metacomment [Meta]	Any comment of the transcriber about any metalinguistic event (e.g. about noises occurring within a pause), to be transcribed as in 5.3 (see Table below)
Comment [Com]	Any comment of the transcriber about the transcription process (e.g. when help is needed to decide between two possible choices for one word)
Status [Status]	Name of the transcriber. Status of the transcription: underway, done, corrected, revised...



1.2. Conventions of transcription

The convention used is mainly inspired by Valibel convention of transcription. However, some changes have been made according to ICOR and Ciel-F convention. Several rules are specific to the CorpAGEst project. The conventions are described in the table below (in French).

Transcription principles

How to calculate duration of pauses?

In the CorpAGEst project, it has been decided to manually annotate the duration of pauses (rather than to have recourse to automatic calculation of pausing), mainly in order to keep the perceptive dimension of the fluency of older speakers' speech, which is even more fluctuant in later life. This not means that an automatic detection of pauses could not be envisaged for further investigation.

What convention for pausing?

It has been decided to time all the pauses that last longer than 200 ms by looking at duration such as they appear in Praat, then transposed by numbers into parentheses. For instance “(2.4)” stands for “2 seconds and 400 milliseconds”. Micro-pauses – that last less than 200 ms – are coded by the “(.)” convention (see the table below).

We chose not to follow the Valibel conventions for pausing (“/” for short pauses and “//” for long pauses), since the same symbol is also used there for the notation of false starts (“/” without any space before) and is thus prone to errors during the transcription process.

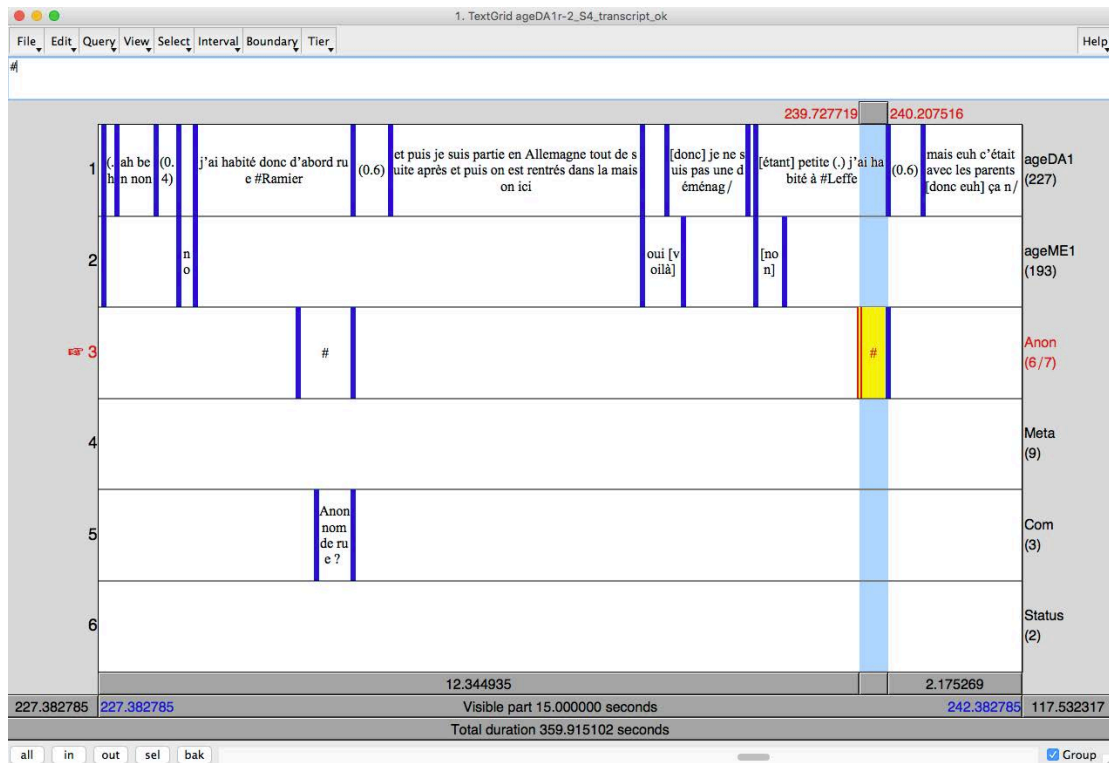
	Phénomène	Convention	Exemple	Correspondance
1	Identification des corpus et locuteurs			
1.1	Code de locuteur	Un code unique est attribué à chaque locuteur (5 lettres suivies d'1 chiffre) : <ul style="list-style-type: none"> Les 3 premières lettres renvoient au nom du corpus (« age » pour les corpus CorpAGEst et CorpAge). Les 2 lettres suivantes correspondent aux initiales réelles du locuteur (rem. : le véritable nom de la personne ne figure nulle part en toutes lettres, ni dans le corpus, ni dans les métadonnées – cf. section sur l'anonymisation). Le dernier chiffre est le chiffre 1 par défaut (avec possibilité d'attribuer les chiffres 2, 3, 4, etc. pour distinguer les homonymes, le cas échéant). 	<i>agePA1</i> pour <i>André Petit</i>	Valibel
1.2	Code d'enregistrement	Un code est attribué à chaque enregistrement. Il s'agit du code du premier locuteur (l'interviewé dans les entretiens semi-dirigés ; la personne prenant la parole en premier lieu dans le cas de conversations spontanées) suivi de la lettre « r » pour <i>recording</i> . Dans le cas où plusieurs enregistrements portent le même nom, on les distingue en ajoutant 2, 3, etc. précédé d'un tiret (par ordre chronologique d'enregistrement).	<i>agePA1r-1</i> <i>agePA1r-2</i>	Valibel, CorpAGEst
2	Orthographe et typographie			
2.1	Orthographe	Respect de l'orthographe conventionnelle. Dans le cas où le locuteur prononce un mot qui n'est pas répertorié dans le dictionnaire, l'orthographe reflètera au mieux la prononciation. Dans le cas d'emprunts à une autre langue, ceux-ci sont transcrits suivant les standards orthographiques de la langue source (rem.: le standard est l'orthographe Feller pour le Wallon).		Valibel, Ciel-F
2.2	Ponctuation	Le point et la virgule ne sont pas utilisés. Le point d'interrogation est noté lorsque l'on observe une montée intonative dans une question à la forme déclarative.	l'endroit idéal ? (<i>ageBN1r-2_sample4</i>)	Valibel, Ciel-F
2.3	Usage de la majuscule	Seuls les noms propres et les noms à référent unique commencent par une majuscule (ainsi que les titres).		Valibel, Ciel-F
2.4	Titres	Le trait d'union et la majuscule sont utilisés dans la transcription des titres	Le-soir	Valibel
2.5	Chiffres	Les chiffres et nombres sont transcrits en toutes lettres et avec des traits d'union	dix-neuf-cent-quatre-vingt-trois	Valibel, Ciel-F, ICOR
2.6	Abréviation	Pas d'utilisation d'abréviation conventionnelle	<i>etcetera</i> et pas <i>etc.</i>	Valibel, Ciel-F
2.7	Sigles et acronymes	Un sigle (prononcé en épelant les lettres) est noté en majuscules sans espace et sans point. Un acronyme (prononcé comme un mot) est noté comme un nom (avec majuscule).	CNRS Setca	Valibel
2.8	Morphologie	<ul style="list-style-type: none"> Accord du participe passé : L'accord au pluriel est toujours noté. En ce qui concerne l'accord au féminin, celui-ci n'est pas noté uniquement dans le cas où sa non-réalisation est audible chez le locuteur 		Valibel

			<ul style="list-style-type: none"> • « Il y a » : Quelle que soit la prononciation, nous notons toujours la forme standard <i>il y a</i> (et pas <i>y'a</i> ou <i>ya</i>) 		
			<ul style="list-style-type: none"> • Négation : « n(e) » : Lorsque la présence ou l'absence de la particule de négation <i>n'</i> n'est pas audible (entre le pronom <i>on</i> et un verbe commençant par une voyelle), on note le <i>n'</i> d'office 	on n'est pas certains d'arriver à temps	
			<ul style="list-style-type: none"> • Variante verbale non standard : Si le locuteur produit une variante verbale non standard, celle-ci est annotée telle que produite. 	j'ai perdu un livre	
	2.9	Interjections, onomatopées et particules discursives	Les interjections, onomatopées et particules discursives sont transcrites selon la liste suivante: ah, ah la la, ah là là, aïe, areu, arf, atchoum, badaboum, baf, bah, bam, bang, bé [be], bè [bɛ], bêêê, ben [be~], beurk, bing, boh, boah, bouh, boum, broum, cataclap, clap clap, coa coa, cocorico, coin coin, crac, croa croa, cuicui, ding, ding deng dong, ding dong, dring, eh [e], eh ben [e be~], eh bien, enfin [a~fe~, fe~], etcetera, euh, euhm, f [f], ff [f:], flic flac, flip flop, froufrou, frou fou, glouglou, glou glou, gnagnagna, groin groin, grr, hein, hep, hi han, hé hé, hip hip hip hourra, hourra, hu, hum (raclement de gorge), m [m], mm [m:], mm mm (acquiescement), mêêê, meuh, mf [mf], mff [mf:], miam, miaou, moah, moh, moui, mouais, m'enfin, of, oh, oh la la, oh là là, ok, ouah, ouah ouah, ouais, ouf, ouh, ouille, oula, ouh là, ouh la la, ouh là là, oup, oups, p [p], paf, pan, patatras, pchhh, pchit, pf [pf], pff [pf:], pfiou, pfou, pfoua, pif-paf, pin pon, pioupiou, plouf, pof, pouet, pouet pouet, pouf, psst, pt, roh, roh la la, roh là là, rohr, ron ron, schlaf, snif, splaf, splatch, sss, t, tacatac, tagada, tchac, teuf teuf, tic tac, toc, tut tut, tss, vlan, vroum, vrrr, wo, wouah, wouaw, wouf, waf, zip.		Valibel, Ciel-F, TCOF, CorpAGEst
3 Hésitations et amorces					
	3.1	Notation exhaustive de la production verbale	Les mots incomplets, les hésitations, les répétitions et les amorces sont transcrits.		
	3.2	Amorces	Dans le cas d'une interruption (avec ou sans reprise ultérieure), le mot amorcé est immédiatement suivi de la barre oblique (sans espace).	on a de plus en pl/ (.) on est de plu/ (ageBN1r-2_sample4)	Valibel
			Le pronom personnel <i>il</i> peut être prononcé [il] ou [i]. Quand il est prononcé [i], même de manière répétitive, on transcrita <i>il</i> .	(prononcé [i i i]): L1 il il il pense	
4 Phénomènes temporels					
	4.1.	Chevauchement	Les chevauchements seront encadrés par des crochets ouvrant et fermant (« [» et «] ») au sein de chaque Tier locuteur. Rem : aucun espace ne doit être inséré entre le crochet et le texte prononcé en chevauchement.	L1 : il faut changer (0.7) quand on sait encore [faut] pas le faire trop tard (0.5) L2 : [ouais] (ageBN1r-2_sample4)	Ciel-F ICOR
			Un évènement non langagier peut également être transcrit en tant que chevauchements (afin de rendre la transcription sous forme de texte plus précise). On veillera cependant à scinder la parole ou l'évènement non langagier afin de distinguer l'empan où il y a chevauchement dans l'empan où il n'y en a pas.	L1 : le lieu idéal pour bien vieillir [[((rire))]] L2 : [oui] (ageBN1r-2_sample4)	

4.2	Silence, pause	Les silences sont chronométrés : la durée est indiquée en secondes entre parenthèses. Les valeurs décimales au centième seront arrondies au dixième supérieur pour les valeurs entre 5 et 9 (par ex. (1.16) devient (1.2)) et au dixième inférieur pour les valeurs entre 1 et 4 (par ex. (1.12) devient (1.1)). On note le 0 après le point, le cas échéant.	(1.2) (2.0)	Ciel-F, ICOR
		Les micro-pauses, c'est-à-dire les silences d'une durée inférieure à 200 ms, sont annotées à l'aide d'un point.	(.)	
4.3	Attribution des silences/pauses	Les silences/pauses ne sont pas attribués. Les pauses sont donc annotées dans la Tier du locuteur qui détenait la parole avant celle-ci. Par conséquent, les pauses de plus de 200 ms n'admettent pas le chevauchement.		CorpAGEst
		En cas de début simultané, la pause est annotée dans la Tier du locuteur qui occupait le canal avant la prise de parole simultanée.		
4.4	Aspiration, prise de souffle	L'aspiration est notée par la lettre ".h" précédée d'un point, si possible dans un intervalle distinct. Elle est attribuée au locuteur qui la produit.	(.h)	Ciel-F, ICOR
4.5	Expiration, soupir	L'expiration est notée par la lettre "h" (sans le point), si possible dans un intervalle distinct. Elle est attribuée au locuteur qui la produit.	(h)	Ciel-F, ICOR
4.6	Ouverture de bouche	Le son dental ou labial émis lors de l'ouverture de bouche (parfois accompagné d'un claquement de langue) est noté par les lettres ".tsk" (dental) ou ".mt" (labial) précédées d'un point, si possible dans un intervalle distinct. Ce son, qui précède généralement la prise de parole, est attribué au locuteur qui le produit.	(.tsk) (.mt)	ICOR, CorpAGEst
5	Production vocale (paraverbale)			
5.1	Production vocale qui accompagne la parole	Le texte concerné par la production vocale est encadré à l'aide de chevrons ("<" et ">") et la description du style vocal est notée entre doubles parenthèses avant la transcription. <((description)) transcription>	L1 : <((rire)) on est> .h on est libre en quelque sorte (ageBN1r-2_sample4)	Ciel-F, ICOR
5.2	Production vocale isolée	Les productions vocales isolées comme le rire ou la toux sont indiqués entre doubles parenthèses dans la Tier du locuteur responsable de la production. Quand elles ne peuvent être attribuées à un locuteur, elles sont notées dans la Tier "Meta".	((rire))	Ciel-F, ICOR
5.3	Orthographe des éléments paraverbaux vocaux	La transcription des éléments paraverbaux respectera l'orthographe ci-dessous: ((bâillement)), ((chantonné)), ((chuchoté)), ((imitation)), ((rire)), ((soupir)), ((toux))		
6.	Multitranscription et passage inaudible			
6.1	Doute	Lorsque le transcripteur doute de sa transcription ou qu'il ne peut se décider sur la forme prononcée (parmi plusieurs possibilités), il note entre accolades la forme qui lui semble être la plus probable.	[ouais {l'âge de la pen/} (.) oui] (ageDA1r-1_sample2)	Valibel
6.2	Passage inaudible	Les parenthèses sont utilisées pour l'indication d'un passage incompréhensible - (x) = une syllabe inaudible - (xx) = un groupe de syllabes inaudible - (xxx) = un passage plus long inaudible	que c'est une période euh (.) oui émotionnellement (x) (.) (ageDA1r-1_sample2)	Valibel

1.3. Text/sound anonymization

- In the speaker's tier: names, surnames and place-names are directly anonymized in the orthographic transcription. The alias is preceded by a “#” symbol.
- In the Anon's tier: the passage that have to be anonymized is precisely (all the passage and only the passage) delimited and pointed by the “#” symbol.
- A script will be used to anonymize the sound on the basis of the Anon's tier (Daniel Hirst's script: hdl:11041/sldr000526)



Anonymization principles

How to choose and pick an alias?

The alias must:

- Begin by the first letter of the original name
- Be close of the original ethnic consonance
- Contain the same number of syllables than the original name
- Be listed on the following websites for French-speaking Belgian names:
 - Family names: <http://www.nom-famille.com/noms-les-plus-portes-par-initiale.html>
 - First names : https://fr.wikipedia.org/wiki/Liste_de_pr%C3%A9noms_fran%C3%A7ais_et_de_la_francophonie#Pr.C3.A9noms_f.C3.A9minins, <http://meilleursprenoms.com/>
 - Placenames : http://fr.wikipedia.org/wiki/Listes_des_villes_du_monde or http://fr.wikipedia.org/wiki/Cat%C3%A9gorie:Village_de_Wallonie

How to avoid mismatching alias with each other?

Most importantly, the alias must be listed with the original name (and possibly the speaker code) in an Excel spreadsheet in order to assign (only once) pseudonyms and alias in a structured and non-redundant manner, taking into account the possibility that the same place or person can be mentioned in distinct recordings.

1.4. Transcription revision

It is strongly recommended to check orthographic transcriptions in a two-step process:

- 1) Revision of the transcription with regard to the sound
- 2) Revision of the transcription without regard to the sound, there are two solutions.

Moreover, a second analyst should reread the transcription in a final step, with recourse to the sound signal.

1.4.1. Revision *with* the sound signal

There are two solutions:

- Reread the Textgrid using Praat (Note that it is possible to navigate quickly from one interval to another using the shortcut ALT+ Keyboard arrows).
- Open the transcription in the software *Transformer 6* (Be careful: this software can bug and is not Mac compatible. So, the first option is preferred).

1.4.2. Revision *without* the sound signal

Reread the transcription without the sound can be helpful to spot spelling mistakes, transcription mistakes, etc. An output format (cf. below, section 3.) can be used to facilitate the revision: preferably the table format which is more pleasant to read.

When the mistakes are spotted, you can change your transcription directly in *Praat* or by opening the file with a text editor (e.g. *Notepad++*). The *Notepad++* solution is quick: you can directly change in the file using the search and replace option.

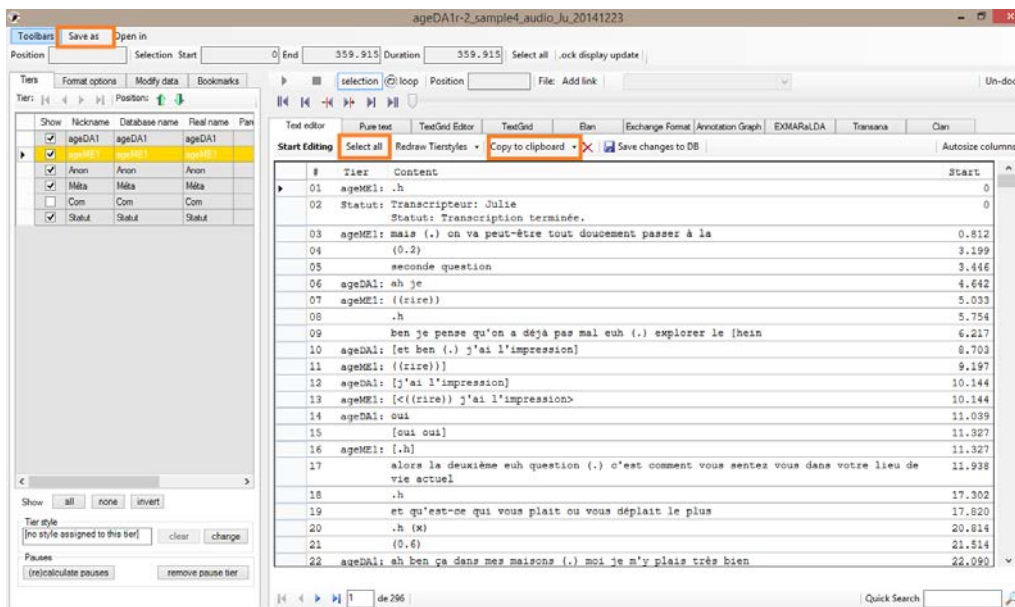
Rem.: Always do backup copies of your files. They can be useful in case of problem.

1.5. Transcription export (*Transformer* program)

Using *Transformer 6*, you can obtain three output formats that are described further (**NOT Mac compatible!!!**).

To proceed, follow the steps:

- 1) Import the textgrid(s): “Upload files” function
- 2) Go to the main window and select the export format needed for an “express” export towards .txt, Praat, ELAN, EXMARaLDA, etc.



1.5.1. Visualization of the raw text output

ageBN1: ((clap))	0
ageFA1: (1.3)	0.346
(.tsk)	1.622
(.h)	1.855
(0.6)	2.513
bien voilà alors je vais tout tout d'abord te poser la	3.131
la question de savoir si tu es toujours d'accord que	
les entretiens soient filmés	
	9.181
– euh et enregistrés	9.427
(1.3)	10.734
ageBN1: (.mt)	12.089
oui je suis d'accord	12.171
(0.5)	12.939
ageFA1: ok	13.433
(0.3)	13.797
(.h)	14.128
alors euh on va discuter un petit peu toutes les deux	14.705
pour explorer le thème de rapport par rapport à l'âge	
(.h)	19.861
et aux différents moments de de ta vie et je te propose	20.272
que nous parlions des grandes étapes de ta vie	
(0.6)	24.593
[(.h)]	25.170

1.5.2. Visualization of the XML output

```
</Metadata>
<Timeline id="Transformer_Timeline1" />
<AG id="Transformer_AG1" type="type" timeline="Transformer_Timeline1">
  <Anchor id="Transformer_AG1_Anchor" offset="1672" unit="milliseconds" />
  <Annotation id="Transformer_AG1_Annotation1" type="ageBN1" start="Transformer_AG1_Anchor"
end="Transformer_AG1_Anchor">
  <Feature name="description">bah on peut peut-être tout doucement passer à à la deuxième [question
(.) euh] (.) de l'entretien</Feature>
</Annotation>
  <Annotation id="Transformer_AG1_Annotation2" type="ageBN1" start="Transformer_AG1_Anchor"
end="Transformer_AG1_Anchor">
  <Feature name="description">[.h]</Feature>
</Annotation>
  <Annotation id="Transformer_AG1_Annotation3" type="ageBN1" start="Transformer_AG1_Anchor"
end="Transformer_AG1_Anchor">
  <Feature name="description">[oui] (.) et je vous dirai immédiatement nous avons une grande maison
(0.3)</Feature>
</Annotation>
  <Annotation id="Transformer_AG1_Annotation4" type="ageBN1" start="Transformer_AG1_Anchor"
end="Transformer_AG1_Anchor">
  <Feature name="description">[.h]</Feature>
</Annotation>
  <Annotation id="Transformer_AG1_Annotation5" type="ageBN1" start="Transformer_AG1_Anchor"
end="Transformer_AG1_Anchor">
  <Feature name="description">maintenant nous sommes retraités (.) je vous ai dit (.) mon mari a quatre-vingt-
deux ans .h</Feature>
```

2. Text-sound alignment

The recommendations are here for the alignment at the level of word unit. Note that one part of the spoken data have also been aligned at the level of phones within the framework of Duboisindien's PhD Thesis.

2.1. Automatic alignment (*EasyAlign* plugin)

Size of the files to run the plugin

It is recommended to compress the audio files first before starting with the alignment procedure. If it has been recorded at 44.100 Hz, it may be the case that you should compress the file at 22.050 Hz (for instance, in Audacity or Praat) in such a way that the plugin could run.

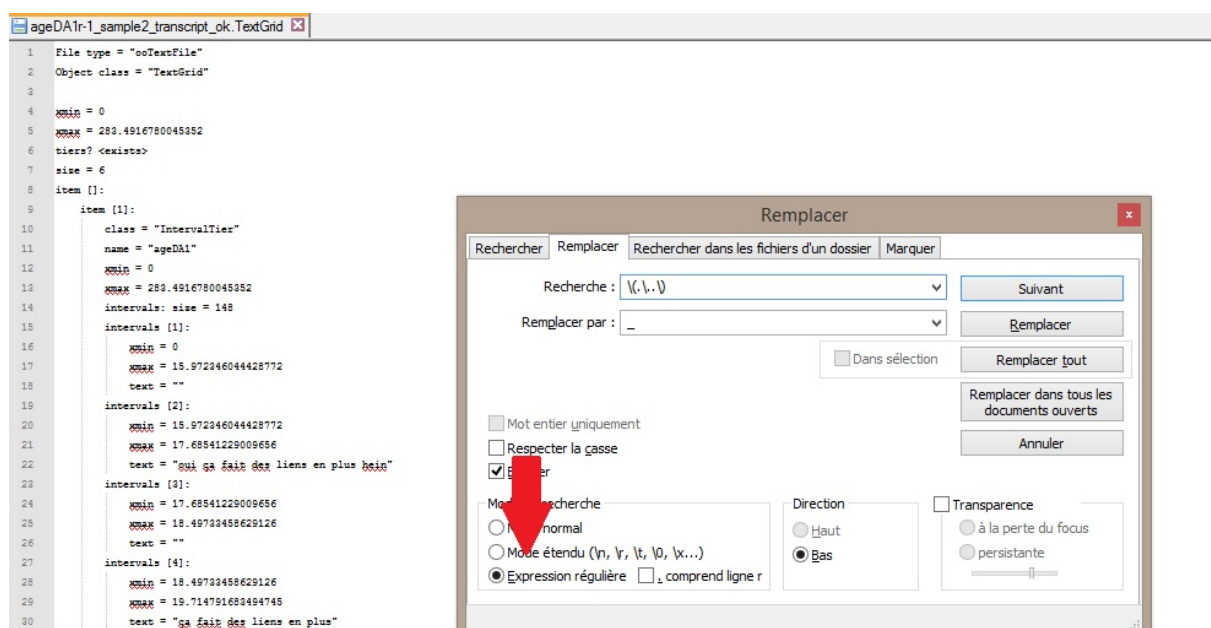
2.1.1. Creation of a "simplified" textgrid

In order to avoid the phonetic transcription of the pauses duration ((1.4), for example) it is recommended to create a new copy of the textgrid with a simplified version of the orthographic transcription.

How to proceed?

- 1) Always copy the initial textgrid
- 2) Open the copy in a text editor such as "Notepad ++", "JEdit"
- 3) Using the search function and delete everything that appear in parenthesis (including pauses duration, short pauses, comments, styles, breath, etc.)
 - a. Select the option which use regular expressions (Regex)
 - b. Search "`\\(\\.|\\)|\\{|\\}|#`" and replace by nothing where "`.*`" means to search every sign that is repeated within a parenthesis. Note that it is also possible to precise the nature of the signs that possibly occur within the parentheses (in order to avoid any deletion of unintended segments), by using this Regex "`\\([a-z,â,ê,à,é,è,ê,î,ô,û,ç,ë,ï,ü\\.,0-9]*\\)`" OR "`\\([a-z,â,ê,à,é,è,ê,î,ô,û,ç,ë,ï,ü\\.,0-9]*\\)`" (depending on the text editor used), which means to search every chain including the mentioned signs, with the exception of spaces, into parenthesis.

Rem.: the "<", ">", "[", "]" symbols cannot be changed at that stage, because used in the language of the Text Editor, then must be manually deleted later in the data treatment.
- 4) Do not forget to save the new textgrid adding the note "simplified"



2.1.2. Isolation of each speaker in one textgrid

In order to facilitate the automatic alignment using the sound from each speaker microphone, it is recommended to isolate each speaker's tier in one textgrid.

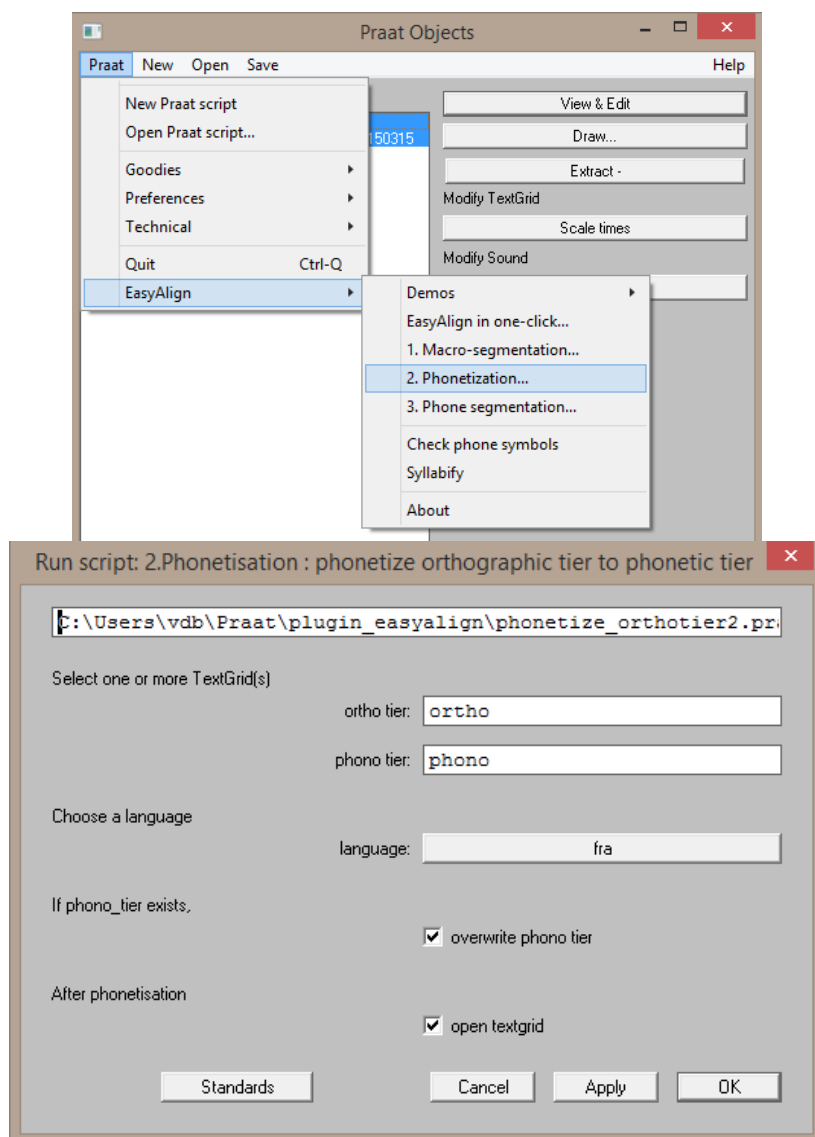
How to proceed?

- 1) In Praat, open the "simplified" version of the textgrid
- 2) Select it and use the function "extract one tier"
- 3) Indicate the number the tier that you want to isolate (usually 1 or 2)
 - ⇒ The textgrid for one speaker is created (textgrid B)

Tip: in order to facilitate the following steps, you can rename the speaker tier "ortho". If you do not do it, you must think, during the following steps, to change the name of the ortho tier in the EasyAlign plugin.

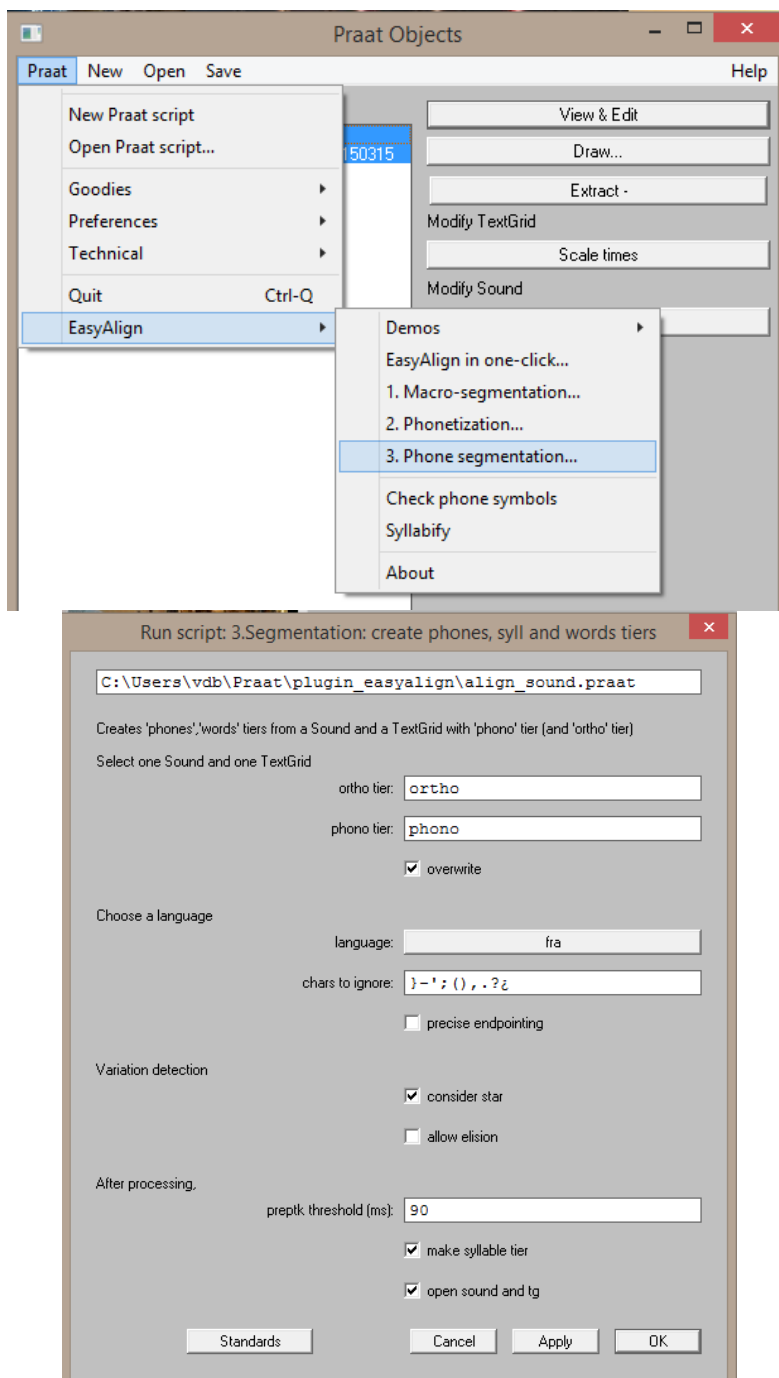
2.1.3. How to use the EasyAlign plugin

- 1) Open the appropriate sound file
- 2) Select the textgrid B and the sound file
- 3) Select EasyAlign > "phonétisation"



Note: the first step ("macrosegmentation") is useless because you have already segmented when you were transcribing.

- 4) Manually, check the phonetization. Suppress the “*”. “*” is used by the script to indicate an optional phoneme. Either the phoneme is pronounced and you can only suppress the “*”, or the phoneme is not pronounced, then suppress the “*” and the preceding letter (if the phoneme is added to the beginning of the word) or the following letter (if the phoneme is added to the end of the word). This step is also the occasion to delete the following symbols from the ortho tier “[“, “[”, “<”, “>”.
- 5) Select EasyAlign > “phonosegmentation”



Note: think about reading the info windows in order to see if there are any problems and why.

- (i) If some intervals are misformatted: check if the number of words in the ortho tier and in the phono tier is the same. Correct it if it is not the case.
- (ii) Some intervals can be too long: try to make them shorter.
- (iii) The phonetic transcription can be wrong: check it and correct it. It especially concerns words such as : *mm, bé...*
- (iv) Redo the step of phonosegmentation

2.2. Manual verification

With regards to the sound, check the alignment: do not hesitate to adjust the boundaries. If some intervals are still not aligned, do it manually.

Tip: The shortcut “Shift + use of the mouse” facilitate the displacement of the frontiers: they will remain perfectly aligned in each tier.

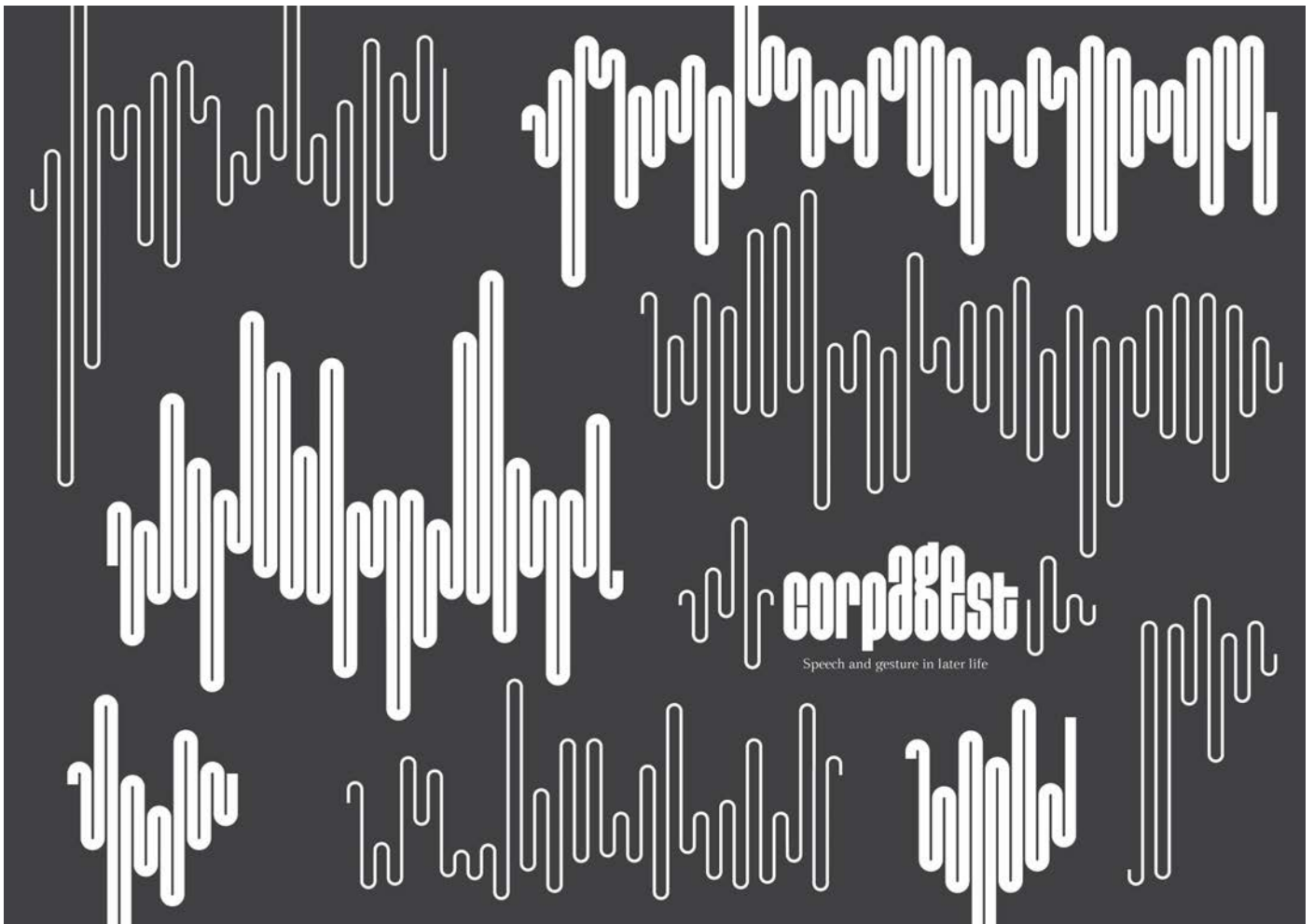
2.2.1. Toward the final textgrid

- 1) Using a text editor replace every “_” by nothing in order to be cleaner when you will import it in Elan.
Tip: Then, you can use the option “Merge similar consecutive intervals” to suppress every empty intervals.
The textgrid will be cleaner.
- 2) In each speaker textgrid, change the name of the word tier : add the speaker code.
- 3) Extract the tiers words of each speakers.
- 4) Select the 3 tiers and the initial textgrid (the one in which the pauses durations are annotated) > merge

Do not forget to save the final textgrid and to rename it.

Note: at the end you are supposed to have three different textgrids by sample.

- 1) The final textgrid containing the initial tiers (described in the beginning of the manual) and the tiers containing the words alignment (imported from the textgrid of each speaker)
- 2) Two textgrids (one by speaker) containing the simplified orthographic transcription, the phonetic transcription, the alignment by word, syllable and phoneme and a “status” tier. This is these Texgrids that will be used if more fine-grained prosodic analysis is needed (for instance, study of the intonation periods by Analor, prosodic contour, etc.). However, note that the phonemes and syllables tiers have not been systematically checked within the CorpAGEst project, because they are not crucial for the purpose of the study (with the exception of * that have been checked and of the adjustment of word frontiers, where relevant – see above).



Original Author: Catherine Bolly
With the collaboration of: Julie Kairet

Contact: catherine.bolly@uclouvain.be; cbolly@uni-koeln.de

Web: <http://corpigest.org>

Logo: © The Shelf Company <http://www.theshelf.fr>