Methodology and software for semi-automatic multi-domain annotations: annotating, exploring and sharing data

Brigitte Bigi

Grégoire de Montcheuil

Roma - September, 14th, 2015

Presenters

Brigitte Bigi

- Researcher at the CNRS, Laboratoire Parole et Langage, Aix-Marseille Université, Aix-en-Provence, France
- Computer Scientist working in the field of corpora and annotations:
  - formalization/constitution of corpora,
  - automatic annotation (mainly at the phonetic level, also at the discourse level),
  - multimodality (annotation, exploration, extraction of annotated data),
  - multilinguality (methods and algorithms).
- Author and developer of SPPAS - Automatic Annotation of Speech

Grégoire de Montcheuil

- Research Engineer at the CNRS - Ortolang, Laboratoire Parole et Langage, Aix-en-Provence, France
- Computer Scientist working in the field of Natural Language Processing:
  - written and spoken French morphosyntactic analisys (MarsaTag)
  - treebank creation (4Couv) and analisys (MarsaGram)
- and in the past:
  - information extraction (resume parser)
  - document classification (medical record, competitive intelligence)
  - word sense disambiguation

Content

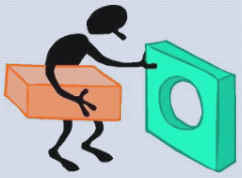
Softwares

Software selection

- The choise of all annotation software must be done carefully, and before the creation of the corpus.
- It is part of the corpus creation framework.

To decide about usefulness and usability, it is necessary to know:

1. about the license,
2. about the ease of use,
3. about the strengths/weaknesses for specific annotation purposes,
4. about the type of data or analysis it is designed for,
5. about its compatibility with other annotated data.


versus


Finding and evaluating appropriate software (1)

1. About the license

- prefer free and open source software:

Even if you can personally afford to pay for a licence for software you may wish to share your methodology with other students or researchers who cannot afford to buy a license.

Finding and evaluating appropriate software (2)

2. About the ease of use

- software or web-services?
- prefer multi-platform software:
  - different scientific communities tend to use Mac OS, Windows or Unix platforms.
  - multi-platform software makes sharing between such communities much easier.

- GUI or CLI usability: prefer usable software!

> If the software requires the help of an engineer each time you need to use it, this will be a serious limitation on your usage.

Finding and evaluating appropriate software (3)

3. About the strengths/weaknesses for specific annotation purposes

- Investigate whether the software has been found to be reliable and is likely to improve the efficiency of workflow, and either accelerate your work or enable you to deal with more extensive data, or both.

Finding and evaluating appropriate software (4)

4. About the type of data or analysis the tool/software is designed for

> When annotating corpora at multiple linguistic levels, annotators may use different expert tools for different phenomena or types of annotation. These tools employ different data models and accompanying approaches to visualization, and they produce different output formats. (Chiarcos et al. 2008)

Finding and evaluating appropriate software (5)

5. About its compatibility with other annotated data

- None of the software are interoperable
- Prefer compatible software

- Estimate the availability to import/export data with a minimum loss of information

Brief overview of some softwares

- Praat
- Elan
- SPPAS

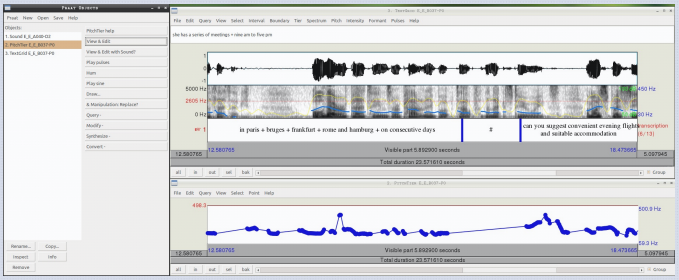
Requirements :

1. free and open-source (GPL),
2. multi-platform, ease of use (GUI), with a tutorial and/or documentation,
3. well-known in their communities, with publications and evaluations.

Praat: design & compatibility

4. Type of data or analysis:
   - Manually annotating sound files
   - Visualizations of audio data: waveform or spectrogram, pitch contour, ...
   - Annotations on multiple layers, called tiers
   - Many plugins for different kinds of analysis
5. Compatibility:
   - Annotation files are in several Praat-specific text formats : Praat-TextGrid
   - Interoperability: none!

Praat: screenshot

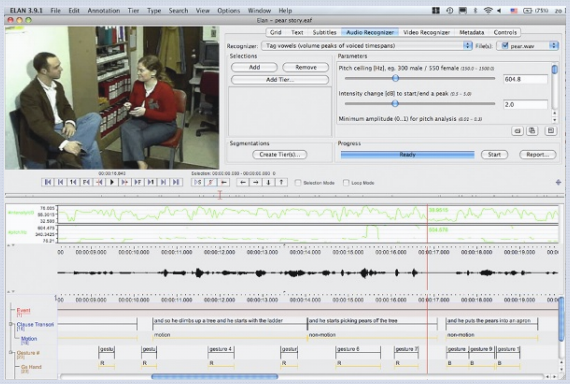


http://www.praat.org

Elan: design & compatibility

4. Type of data or analysis:
   - Creation of complex annotations in video (and audio) resources
   - Annotations can be created on multiple layers, that can be hierarchically interconnected and can correspond to different levels of linguistic analysis
5. Compatibility:
   - Annotation files are in a specific XML format
   - Import from/export to a variety of other formats, including Praat-TextGrid

Elan: screenshot



https://tla.mpi.nl/tools/tla-tools/elan/

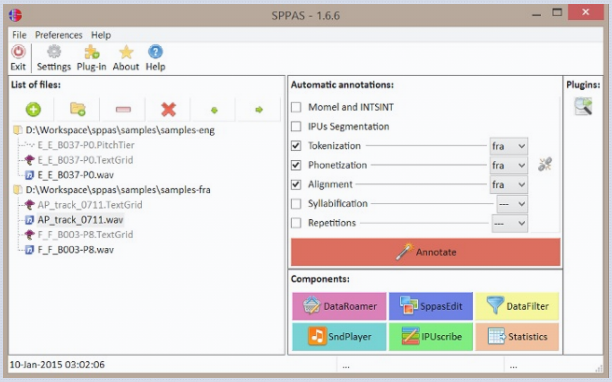
SPPAS: design & compatibility

4. Type of data or analysis:
   - Create, visualize and search annotations for audio data
   - Automatically speech segmentation annotations from a recorded speech sound and its transcription.
   - Plugins, CLI, scripting
5. Compatibility:
   - Annotation files are in a specific XML format
   - Import from/export to a variety of other formats, including: Praat (TextGrid, PitchTier, IntensityTier), Elan (eaf), Annotation Pro (antx), Phonedit (mrk), HTK (lab, mlf), Sclite (ctm, stm), subtitles (sub, srt) Transcriber (trs, import only), Anvil (anvil, import only), CSV

SPPAS: dedicated to automatic annotations

- Language-independent algorithms:
  - language-dependent resources
  - easy and fast to add a new language
- Fully-automatic or semi-automatic (with a procedure outcome report)
- Designed to be able to deal with spontaneous speech
- No limit of the corpus size

SPPAS: screenshot



http://sldr.org/sldr000800/preview/

Summary

- Softwares
  - Selection
  - Examples : Praat, Elan, SPPAS
- An annotation workflow
  - Recording
  - IPU & Transcription
  - Phonetics: tokens, phonemes, syllables
  - Morphosyntax: POS, lemma, chunk
  - Discourse: repetitions
  - Prosody: Momel & INTSINT
  - Gestures
- Exploring
- Sharing
  - Why ? How ?
  - Data Repositories
  - Metadata

An annotation workflow

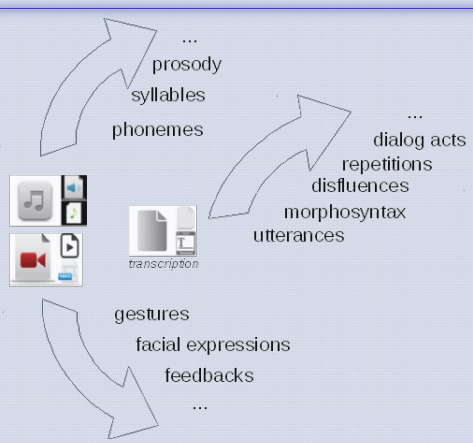
Information

- The proposed methodology is designed for the human analyst (mostly researchers in Linguistics).
- Therefore, we assume that the methodology is general enough to be useful for broad class of research applications.
- Different analytical domains - e.g. speech and gesture - and theoretical perspective require a rigorous organization of the annotation procedure.

Which annotations (in general)?

> A very large number of dimensions have been annotated in the past on mono and multimodal corpora. To quote only a few, some frequent speech or language based annotations are speech transcript, segmentation into words, utterances, turns, or topical episodes, labeling of dialogue acts, and summaries; among video-based ones are gesture, posture, facial expression [...]. (Popescu-Belis, 2010)

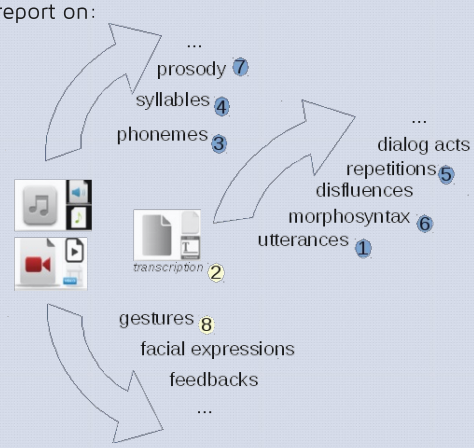
Which annotations (in general)?



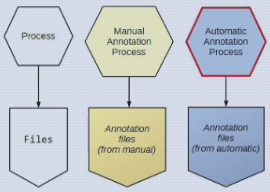
Which annotations (in this tutorial)?

In this tutorial, we will report on:

1. IPUs segmentation
2. Speech transcript (manual)
3. Phonemes and words segmentation
4. Syllables segmentation
5. Repetitions detection
6. Morpho-syntax
7. Momel and INTSINT
8. Gestures (manual)

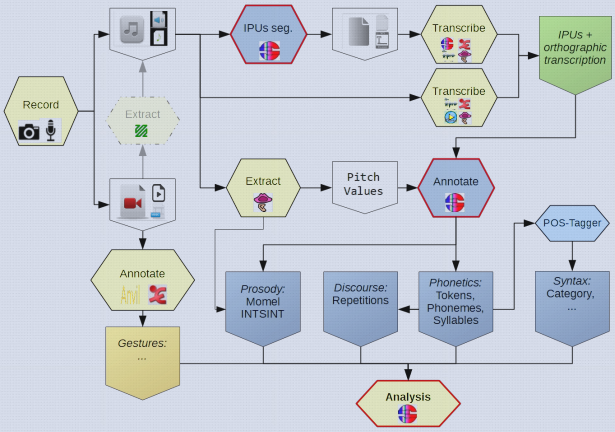


The annotation workflow: legend



Legend of the annotation workflow

The annotation workflow

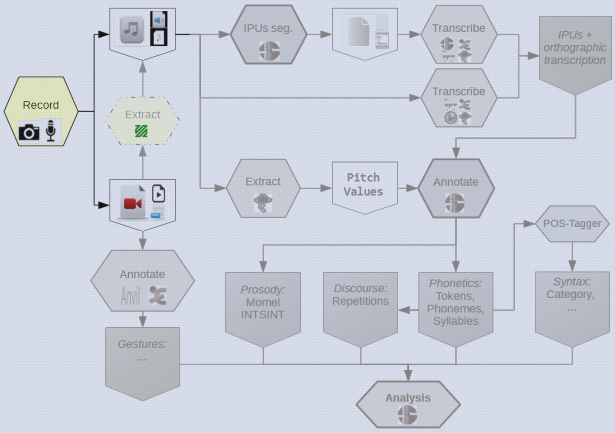


The annotation workflow

The main principle is...

> Garbage in, Garbage out.

Record

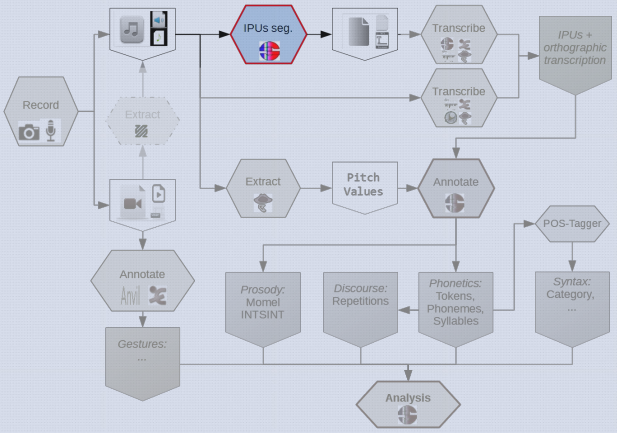


Capturing and recording multimodal data

> The capture of multimodal corpora requires complex settings such as instrumented lecture and meetings rooms, containing capture devices for each of the modalities that are intended to be recorded, but also, most challengingly, requiring hardware and software for digitizing and synchronizing the acquired signals. (Popescu-Belis, 2010)

- Some advices:
  - Audio: avoid noice; one channel per speaker; uncompressed; 16000Hz (or multiple) enough for automatic speech tools
  - Video: prefer H.264 (standard); test the annotation software(s); avoid conversions (record directly into the expected format)
  - Synchronization: use a regular "clap"

IPUs Segmentation



IPUs Segmentation: definition

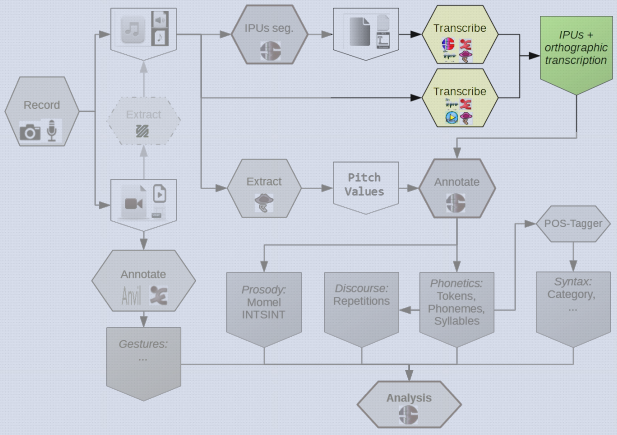
- Automatic segmentation in Inter-Pausal Units
  - is also called Silence/Speech segmentation
- Parameters to define manually:
  - fix the minimum silence duration
  - fix the minimum speech duration
  - both values depend on:
    - the language
    - the speech style
- As results:
  - speech and silences are time-aligned and annotated automatically

IPUs Segmentation: software

- SPPAS is recommended
- A manual verification is recommended

Example of IPUs segmentation: Silences are annotated with # and speech intervals are filled with ipu number

Orthographic Transcription



Orthographic Transcription

- An orthographic transcription is the minimum requirement for a speech corpus,
  - a better representation of pronunciation may be desired for most of research questions
- Orthographic transcription is at the top of the annotation procedure:
  - and remember: "Garbage in, Garbage out."
- Orthographic transcription of spoken language presents considerable challenges.

Orthographic Transcription

- Speech may be annotated for:
  - phonemic transcription;
  - phonetic transcription taking into account details of pronunciation
    - allows a time-alignment at the phoneme-level
    - which is extended to time-alignment at word-level and syllable-level.
  - syntax analysis

Orthographic Transcription

- The better orthographic transcription implies:
  - the better phonetic transcription,
  - thus, the better time-alignment of phonemes,
  - thus, the better time-alignment of tokens,
  - thus, the better syllabification,
  - and so on...
- But, what is "the better" orthographic transcription?
  1. it's a representation of what is "perceived" in the signal
  2. it follows the convention the automatic system is requiring

Orthographic Transcription for spontaneous speech

- One of the characteristics of Spontaneous Speech is an important gap between a word's phonological form and its phonetic realizations.
- Specific realizations due to elision or reduction processes are frequent in spontaneous data.
- It also presents other types of phenomena such as:
  - non-standard elisions,
  - substitutions or addition of phonemes
  - noises, laughter, ...
- All of them intervene in the automatic system

Enriched Orthographic Transcription

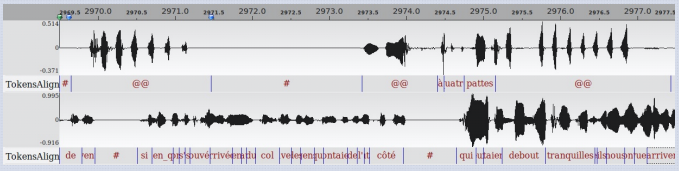
- In speech (particularly in spontaneous speech), many phonetic variations occur:
  - Some of these phonologically known variants are predictable

Transcription:	I	never	get	to	sleep	on	the	airplane
Phonetization:	ay	n.eh.v.e.r	g.eh.t g.ih.t	t.uw t.ix t.ax	s.l.iy.p	aa.n ao.n	dh.ax dh.ah dh.iy	eh.r.p.l.ey.n

  - but many others are still unpredictable (especially invented words, regional words or words borrowed from another language)
- The orthographic transcription must be enriched:
  - it must be a representation of what is "perceived" in the signal.

Enriched Orthographic Transcription

- In speech (particularly in spontaneous speech), many kind of events can occur like breathes, laughter, ...



Enriched Orthographic Transcription

- An EOT must include, at least:
  - Filled pauses
  - Short pauses
  - Repeats
  - Truncated words
  - Noises
  - Laughter
- An EOT must also include:
  - un-regular elisions
  - specific pronunciations
- An EOT may include:
  - all elisions

Enriched Orthographic Transcription: convention

- Any EOT must follow a convention
- The EOT is the input for automatic systems... and the transcription convention depends on the tool/software.
- So... you must read the documentation before starting to transcribe!

Train you first to transcribe and to use the annotation software!

SPPAS transcription convention

- truncated words, noted as a '-' at the end of the token string (an ex- example)
- noises, noted by a '*'
- laughs, noted by a '@'
- short pauses, noted by a '+'
- elisions, mentioned in parenthesis
- specific pronunciations, noted with brackets [example,eczap]
- comments are noted inside braces or brackets without using comma {this} or [this and this]
- liaisons, noted between '=' (an =n= example)
- morphological variants with <like,lie ok>
- proper name annotation, like $John S. Doe$

Transcription example 1 (Conversational speech)

- EOT:

> donc + i- i(l) prend la è- recette et tout bon i(l) vé- i(l) dit bon [okay, k]

- derived Standard orthograph:
  - donc il prend la recette et tout bon il dit bon okay
- derived Faked orthograph:
  - donc + i i prend la è recette et tout bon i vé i dit bon k

Transcription example 2 (Conversational speech)

- EOT:

> ah mais justement c'était pour vous vendre bla bla bla bl(a) le mec i(l) te l'a emboucané + en plus i(l) lu(i) a [acheté,acheuté] le truc et le mec il est parti j(e) dis putain le mec i(l) voulait

- Standard orthograph:
  - ah mais justement c'était pour vous vendre bla bla bla bla le mec il te l'a emboucané en plus il lui a acheté le truc et le mec il est parti je dis putain le mec il voulait
- Faked orthograph
  - ah mais justement c'était pour vous vendre bla bla bla bl le mec i te l'a emboucané + en plus i lu a acheuté le truc et le mec il est parti j dis putain le mec i voulait

Transcription example 3 (Grenelle II)

- EOT:

> euh les apiculteurs + et notamment b- on n(e) sait pas très bien + quelle est la cause de mortalité des abeilles m(ais) enfin il y a quand même + euh peut-êt(r)e des attaques systémiques

- Standard orthograph:
  - les apiculteurs et notamment on ne sait pas très bien quelle est la cause de mortalité des abeilles mais enfin il y a quand même peut-être des attaques systémiques
- Faked orthograph:
  - euh les apiculteurs + et notamment b on n sait pas très bien + quelle est la cause de mortalité des abeilles m enfin il y a quand même + euh peut-ête des attaques systémiques

Enriched Orthographic Transcription of 3 corpora

MARC-Fr: Manually phonetized and aligned French Corpus

	CID	AixOx	Grenelle
Duration	143s	137s	134s
Number of speakers	12	4	1
Number of phonemes	1876	1744	1781
Number of tokens	1269	1059	550
Silent pauses	10	23	28
Filled pauses	21	0	5
Noises (breathes,...)	0	8	0
Laughter	4	0	0
Truncated words	6	2	1
Optional liaisons	4	2	5
Elisions (non stds)	60	21	34
Special Pron.	58	37	23

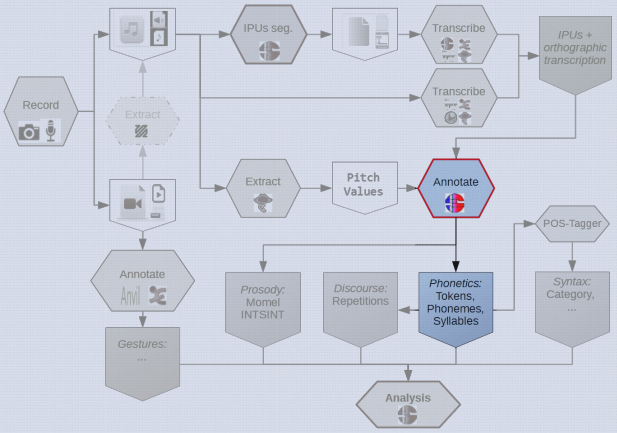
http://sldr.org/sldr000786

Orthographic Transcription... to sum up

- An Enriched Orthographic Transcription is required
- The EOT of a corpus must follow a transcription convention
- Manual Standard orthographic transcription takes 15-20 minutes / minute of speech.
- Manual Enriched orthographic transcription takes 30-45 minutes / minute of speech.

> The automatic systems must be adapted to deal with EOT

Phonemes/Tokens time-alignment



Phonemes and Tokens time-alignment

- A problem divided into 3 sub-tasks:
  1. tokenization
     - text normalization, word segmentation
  2. phonetization
     - grapheme to phoneme conversion
  3. alignment
     - speech segmentation

Tokenization

- Tokenization is also known as "Text Normalization".
- Tokenization is the process of segmenting a text into tokens.
- In principle, any system that deals with unrestricted text need the text to be normalized.
- Automatic text normalization is mostly dedicated to written text, in the NLP community

Tokenization in SPPAS

The main steps of the text normalization proposed in SPPAS are:

- Remove punctuation
- Lower the text
- Convert numbers to their written form
- Replace symbols by their written form (like %, °, ...)
- Word segmentation
  - based on a lexicon.

Tokenization in SPPAS

- From an EOT, SPPAS produces 2 outputs:
  - standard: the text normalization of the standard transcription,
  - faked: the test normalization of the faked transcription.
- Example:

This is + hum... an enrich(ed) transcription {loud} number 1!

- standard: this is hum an enriched transcription number one
- faked: this is + hum an enrich transcription number one

(Bigi 2011)

Phonetization

- Phonetization is also known as grapheme-phoneme conversion
- Phonetization is the process of representing sounds with phonetic signs.
- Phonetic transcription of text is an indispensable component of text-to-speech (TTS) systems and is used in acoustic modeling for automatic speech recognition (ASR) and other natural language processing applications.

> Converting from written text into actual sounds, for any language, cause several problems that have their origins in the relative lack of correspondence between the spelling of the lexical items and their sound contents.

Phonetization in SPPAS

- SPPAS implements a dictionary based-solution
  - consists in storing a maximum of phonological knowledge in a lexicon.
  - In this sense, this approach is language-independent.
- The phonetization process is the equivalent of a sequence of dictionary look-ups
- SPPAS implements a language-independent algorithm to phonetize unknown words.

(Bigi 2013)

Phonetization in SPPAS

By convention, spaces separate words, dots separate phones and pipes separate phonetic variants of a word. For example, the transcription utterance:

Input

the flight was twelve hours long and we really got bored

Output

dh.ax|dh.ah|dh.iy f.l.ay.t w.aa.z|w.ah.z|w.ax.z|w.ao.z t.w.eh.l.v
aw.er.z|aw.r.z l.ao.ng ae.n.d|ax.n.d w.iy r.ih.l.iy|r.iy.l.iy g.aa.t
b.ao.r.d

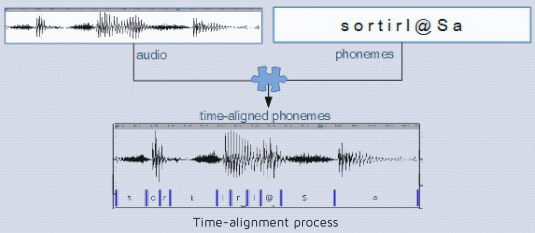
Impact of the Orthographic Transcription on automatic phonetization

- In (Bigi et al. 2012), we compared 3 types of OT:
  1. Standard orthographic transcription.
  2. Enriched 1: Std-OT + short pauses, various noises, laughter, filled pauses, truncated words, repeats.
  3. Enriched 2: Enriched 1 + elisions, particular pronunciations and unusual liaisons.
- Evaluations compare a reference phonetized manually to phonetizations obtained with SPPAS

	Sub %	Del %	Ins %	Err %
CID				
TO standard	2,8	4,5	10,0	17,3
TO enrichie - 1	2,7	1,4	10,3	14,4
TO enrichie - 2	1,8	1,3	3,4	6,5
AixOx				
TO standard	1,4	5,0	3,0	9,5
TO enrichie - 1	1,4	2,3	2,9	6,5
TO enrichie - 2	1,3	1,8	2,5	5,6
Grenelle				
TO standard	1,1	2,8	4,1	8.0
TO enrichie - 1	1,0	1,2	4,1	6,3
TO enrichie - 2	1,3	1,0	1,7	4,0

Alignment

- Alignment is also called phonetic segmentation
- The alignment problem consists in a time-matching between a given speech unit along with a phonetic representation of the unit.



Time-alignment process

> Manual alignment has been reported to take between 11 and 30 seconds per phoneme. (Leung and Zue, 1984)

How to perform Speech Segm. ?

1. Many freely available tool boxes, i.e. Speech Recognition Engines that can perform Speech Segmentation
   - HTK - Hidden Markov Model Toolkit
   - CMU Sphinx
   - Open Source Large Vocabulary CSR Engine Julius
2. Wrappers for such tool boxes:
   - Prosodylab-Aligner: python+HTK
   - P2FA: python+HTK
3. Web-services:
   - WebMAUS
   - Train&Align

How to perform Speech Segm. ?

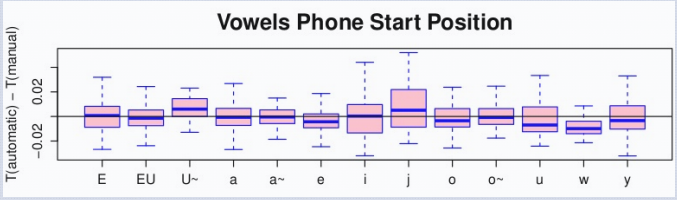
4. Packaged software

- user-friendly,
- with Graphical User Interface,
- with Command-line Interface,
- documentated,
- maintained,
- open-source,
- etc...

SPPAS (python+Julius), available for English, French, Italian, Spanish, Catalan, Polish, Japanese, Mandarin Chinese, Taiwanese, Cantonese

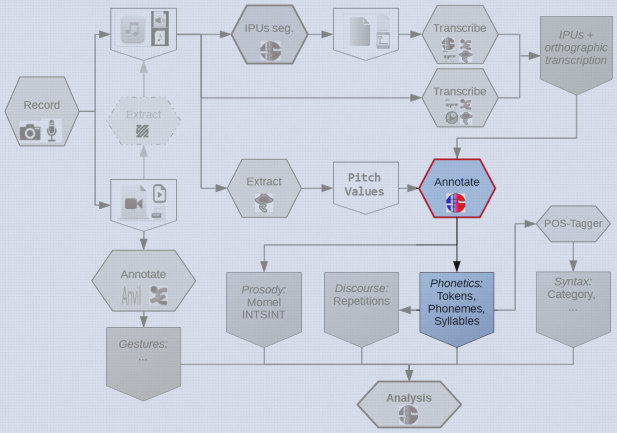
Alignment results in SPPAS

- In average, automatic speech segmentation of French is 95% of times within 40ms compared to the manual segmentation (SPPAS 1.5, September 2014):
  - tested on read speech
  - tested on conversational speech



Results on vowels of French conversational speech

Syllables segmentation



Syllabification by SPPAS

- Automatic annotation
- A rule-based system
- Rules available for:
  - French
  - Italian
- This phoneme-to-syllable segmentation system is based on 2 main principles:
  1. a syllable contains a vowel, and only one;
  2. a pause is a syllable boundary.

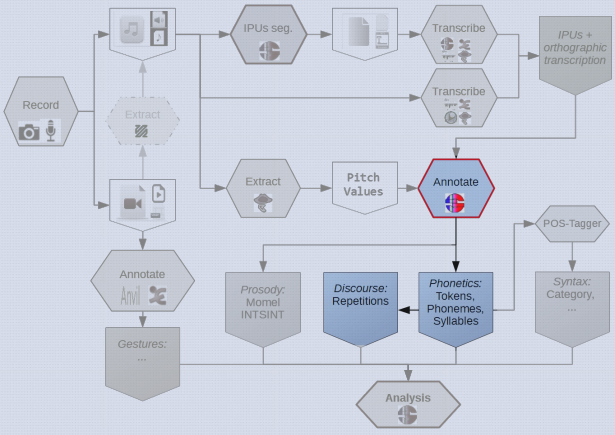
(Bigi et al. 2010)

Syllabification by SPPAS

- Phonemes are grouped into classes, for both French and Italian:
  - V - Vowels,
  - G - Glides,
  - L - Liquids,
  - O - Occlusives,
  - F - Fricatives,
  - N - Nasals.
- Fix rules to find the boundaries between two vowels

Transcription	il expliquait pas vraiment ce qu'il y avait dedans
Phonemes	i l e k s p l i k e p a v r e m ɑ̃ s k i j a v e d ɑ̃
Classes	V G V O F O L V O V O V F L V N V F O V V V F V O V
Syllables Auto	i . lek . spli . ke . pa . vre . mɑ̃ . ski . ja . ve . dɑ̃
Syllables Expert1	i . lek . spli . ke . pa . vre . mɑ̃ . ski . ja . ve . dɑ̃
Syllables Expert2	i . leks . pli . ke . pa . vre . mɑ̃ . ski . ja . ve . dɑ̃

Repetitions detection



Repetitions

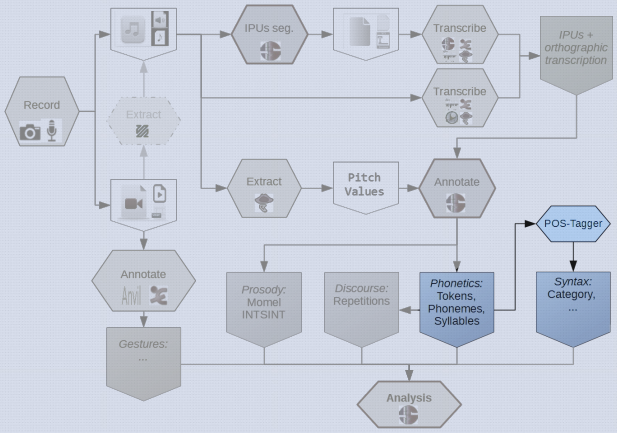
- Other-repetition is a device involving the reproduction by a speaker of what another speaker has just said.
- Other-repetition has been identified as an important mechanism in face-to-face conversation through their discursive or communicative functions

(Bigi et al. 2014)

Repetitions

- Semi-automatic annotation performed by SPPAS
- SPPAS implements:
  - self-repetitions,
  - other-repetitions detection (CLI only).
- The system is based only on lexical criteria, from the time-aligned tokens (or lemmas)
- The system was used to propose a lexical characterization of OR: various statistics was estimated on the detected OR

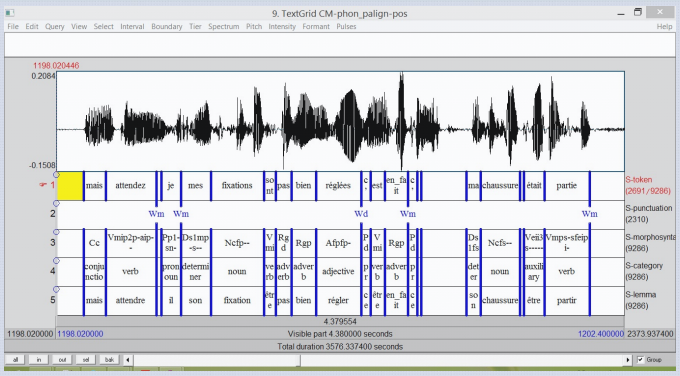
Morpho-syntax



Morpho-syntax

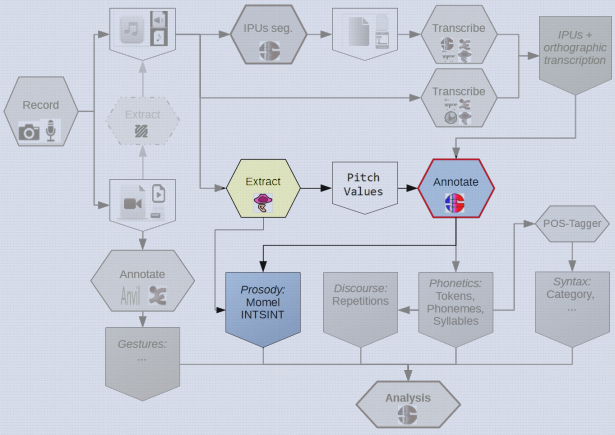
- It is mostly dedicated to written text, in the NLP community
- A system must be adapted to deal with speech, particularly for conversational speech:
  - spoken data are time-aligned and we expect to get a time-aligned morpho-syntax!
  - the lexicon and the probabilities of tokens are different between written texts and speech, so they must be updated.
- At LPL, Stéphane Rauzy and G. de Montcheuil are proposing MarsaTag, for French:
  - http://sldr.org/sldr000841

Example of Morpho-syntax in CID



Example of time-aligned morpho-syntax on conversational speech

Momel and INTSINT

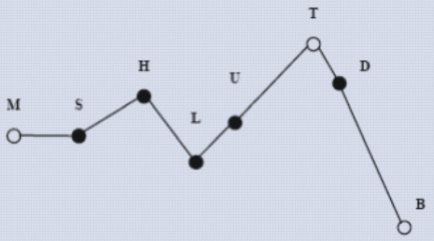


Momel and INTSINT

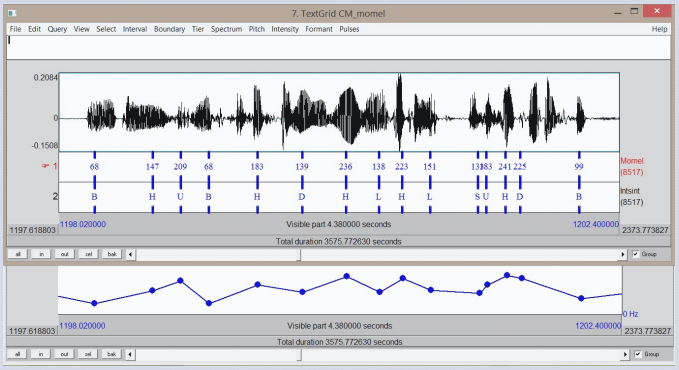
- Momel (modelling melody)
  - algorithm modelling raw fundamental frequency curves with a quadratic spline function
  - target F_0 Points
- INTSINT: an INternational Transcription System for INTonation
  - based on an inventory of minimal pitch contrasts found in published descriptions of intonation patterns
  - surface phonological structure
  - mapping from Momel target points to INTSINT tones

INTSINT

- Absolute tones: T(op) M(id) B(ottom)
- Relative tones: H(igher) S(ame) L(ower)
- Iterative relative tones: U(pstepped) D(ownstepped)



Example of Momel and INTSINT

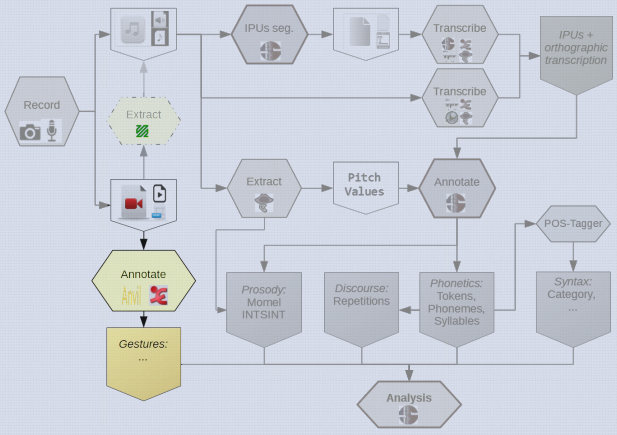


Momel and INTSINT: software

- Momel and INTSINT are available:
  - as a Praat plugin, developped by Daniel Hirst
  - in SPPAS, developped by Brigitte Bigi

(Hirst and Espesser, 1993)

Gestures



Gestures

- What?
  - shape: up/down/sideways/complex/..., single hand/both hands
  - phase: preparation/stroke/post-stroke hold/retraction/...
  - function: deictic/iconic/symbolic/feedback/...
  - ...

(Tellier 2014)

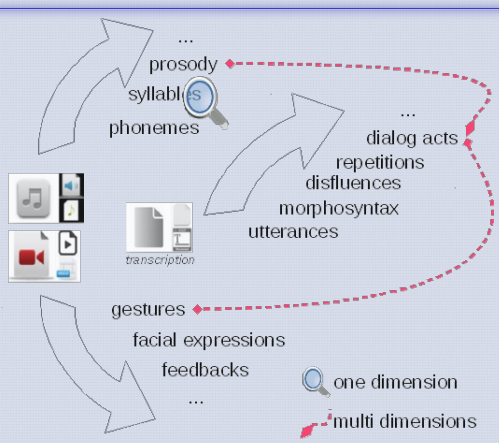
- Use/build an annotation scheme, adapted to the particular framework of your research
- Define the annotation guide, unambiguous
- Validation: intercoder agreement, ...

Summary

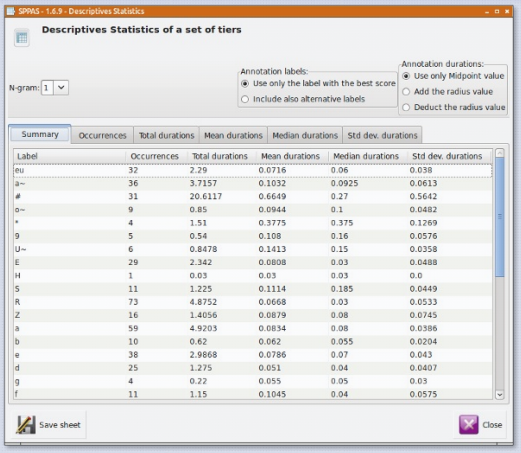
- Softwares
  - Selection
  - Examples : Praat, Elan, SPPAS
- An annotation workflow
  - Recording
  - IPU & Transcription
  - Phonetics: tokens, phonemes, syllables
  - Morphosyntax: POS, lemma, chunk
  - Discourse: repetitions
  - Prosody: Momel & INTSINT
  - Gestures
- Exploring
- Sharing
  - Why ? How ?
  - Data Repositories
  - Metadata

Exploring

Exploring: What ?



Statistics

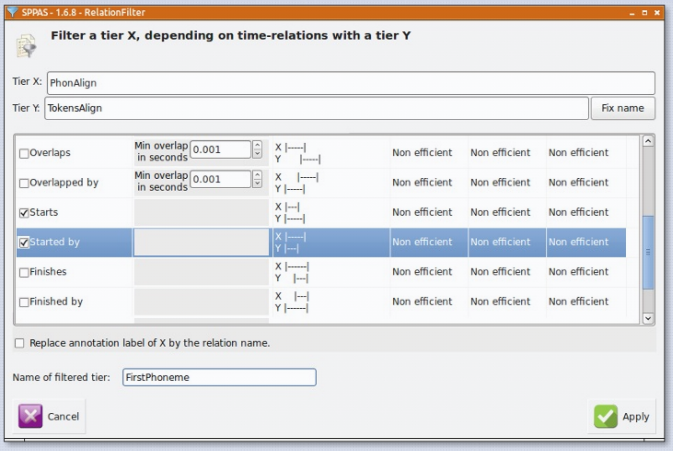


Time relations

Allen's basic temporal relations

Relation		Converse
precedes	a b	b a preceded by
meets	a b	b a met by
overlaps	a b	b a overlapped by
finished by	a b	b a finishes
contains	a b	b a during
starts	a b	b a started by
equals	a b	

Allen in SPPAS



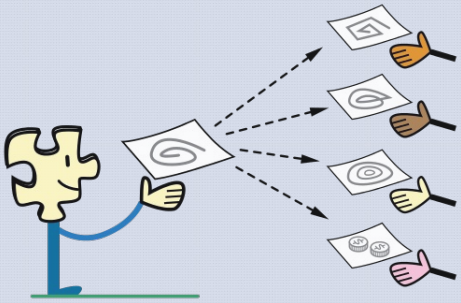
select the first phoneme of each token

Summary

- Softwares
  - Selection
  - Examples : Praat, Elan, SPPAS
- An annotation workflow
  - Recording
  - IPU & Transcription
  - Phonetics: tokens, phonemes, syllables
  - Morphosyntax: POS, lemma, chunk
  - Discourse: repetitions
  - Prosody: Momel & INTSINT
  - Gestures
- Exploring
- Sharing
  - Why ? How ?
  - Data Repositories
  - Metadata

Sharing

Sharing: Why ?



Spread knowledge around the world by making it accessible and reusable

Sharing: How ?

- , , , ...
  - very limited
- Personal / University website
  - perenity ? access rights ? ...
- Data Repository

Some Data Repositories features

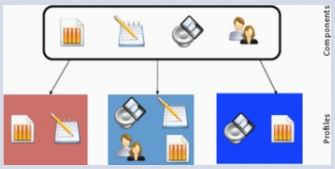
- Store your data
  - secure, mid-/long-term archiving
- Manage access rights
  - public or restricted
- Use metadata
  - classify, search and retreive
  - harvestable => more visibility
- Persistent ID: Handle, DOI, ...
  - easier to cite, reference, reuse

Which Data Repository ?

- Generic? , , ...
  or Specific?
  - LR: The Language Archive , , ...
  - Speech: Bavarian Archive for Speech Signals (BAS), , SLDR/Ortolang , ...
- Criteria:
  - Metadata stored? searchable?
  - Harvested by other archives portals? Open Language Archives Community (OLAC) , (CLARIN),
  - Curators?

Metada ... in a few words

- A lot of acronyms:
  - DCMI: Dublin Core Metadata Initiative => the basis
  - TEI: Text Encoding Initiative => text corpus
  - OLAC: Open Language Archives Community, IMDI: ISLE Meta Data Initiative => multi-media and multi-modal language resources
  - CMDI: Component MetaData Infrastructure => describe and reuse metadata "components" grouped into a "profile"



- Some "components":
  - Session: Recording, Speakers, Languages, Task, ...
  - Annotation: Type, Tool, Tagset, Manual, ...
  - Research: Publication, Project, Institution, Funder, ...

Summary

References

Bigi, Brigitte, Christine Meunier, Irina Nesterenko and Roxane Bertrand (2010). Automatic detection of syllable boundaries in spontaneous speech. Language Resource and Evaluation Conference, pages 3285-3292, La Valetta, Malte.

Bigi, Brigitte (2012). SPPAS: a tool for the phonetic segmentations of Speech. The eight international conference on Language Resources and Evaluation, Istanbul (Turkey), pages 1748-1755, ISBN 978-2-9517408-7-7.

Bigi, Brigitte and Daniel Hirst (2012). SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. Speech Prosody, Tongji University Press, ISBN 978-7-5608-4869-3, pages 19-22, Shanghai (China).

Bigi, Brigitte (2013). A phonetization approach for the forced-alignment task. 3rd Less-Resourced Languages workshop, 6th Language & Technology Conference, Poznan (Poland).

Bigi, Brigitte (2014). A Multilingual Text Normalization Approach. Human Language Technologies Challenges for Computer Science and Linguistics. LNAI 8387, Springer, Heidelberg. ISBN: 978-3-319-14120-6. Pages 515-526.

Bigi, Brigitte, Roxane Bertrand and Mathilde Guardiola (2014). Automatic detection of other-repetition occurrences: application to French conversational speech. 9th International conference on Language Resources and Evaluation (LREC), Reykjavik (Iceland), pages 2648-2652. ISBN: 978-2-9517408-8-4.

Chiarcos, Christian, Stefanie Dipper, Michael Götze, Ulf Leser, Anke Lüdeling, Julia Ritz, and Manfred Stede (2008). A flexible framework for integrating annotations from different tools and tagsets. Traitement Automatique des Langues 49, no. 2: 271-293.

Hirst, Daniel and Robert Espesser (1993). Automatic Modelling Of Fundamental Frequency Using A Quadratic Spline Function. Travaux de l'Institut de Phonétique d'Aix, pages 75-85, vol. 85.

Leung HC and Zue VW (1984). A Procedure for Automatic Alignment of Phonetic Transcriptions with Continuous Speech. ICASSP '84; San Diego, CA. pp. 2.7.1–2.7.4.

Popescu-Belis, Andrei (2010). Managing Multimodal Data, Metadata and Annotations: Challenges and Solutions. Chapter 11, pages 187-199, IDIAP.

Rauzy, Stéphane, Grégoire De Montcheuil and Philippe Blache (2014). MarsaTag, a tagger for French written texts and speech transcriptions. Second Asia Pacific Corpus Linguistics Conference, Hong Kong.

Tellier, Marion (2014). Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés. Discours. Revue de linguistique, psycholinguistique et informatique, vol. 15.