Réunion C3I – 23 février 2012

Brigitte Bigi

Phonetizaton   Automatic   Speech
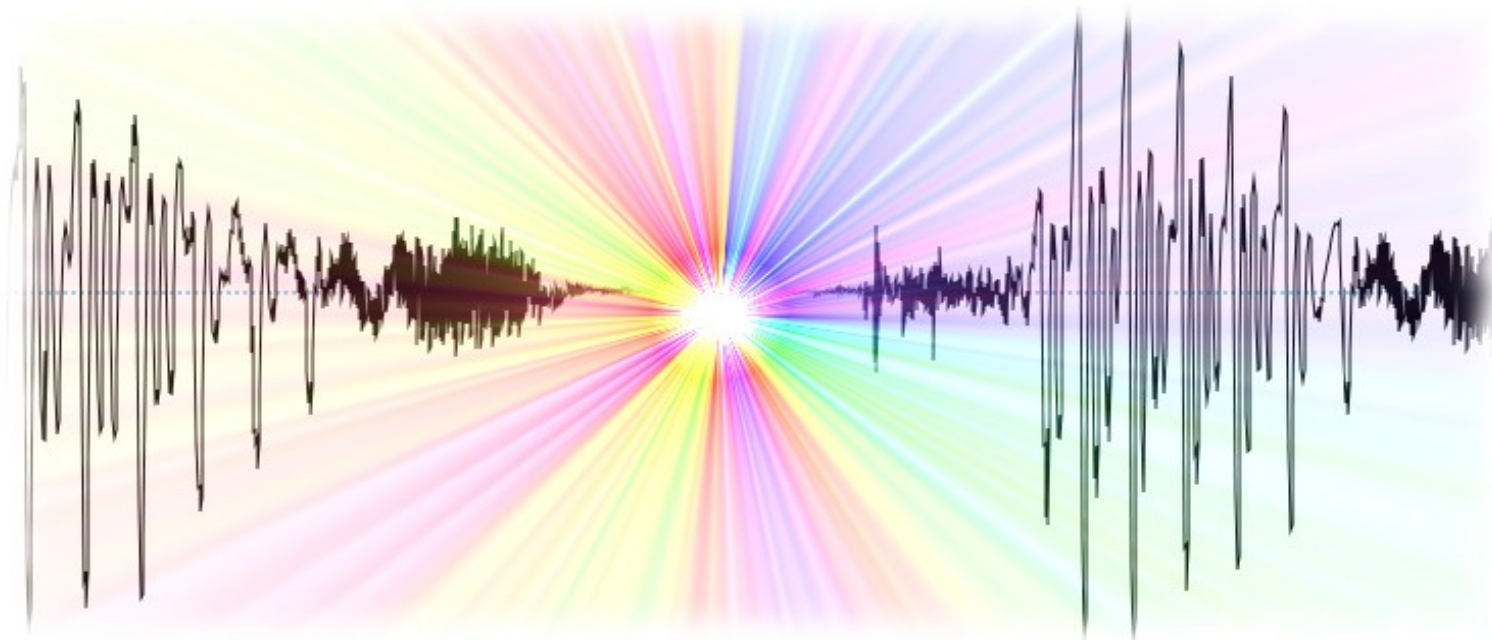   Syllabification   Segmentation
     Alignment   Prosody

SPPAS

Main description

- A tool to produce automatically annotations which includes utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription

- Language-independent

- Currently designed for French, English, Italian and Chinese and there is an easy way to add other languages

GNU Public License

What SPPAS can do today?

- **IPU-segmentation**: utterance level segmentation

- **Phonetization**: grapheme to phoneme conversion

- **Alignment**: phonetic segmentation

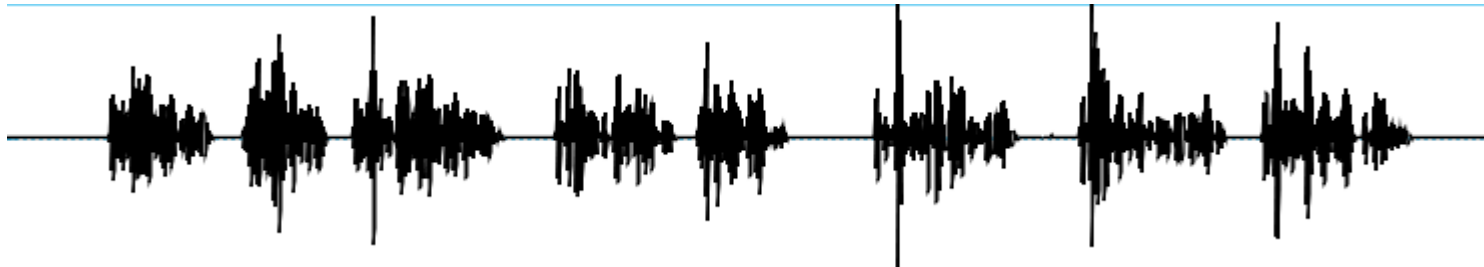- **Syllabification**: group phonemes into syllables

Operating systems:

SPPAS inputs

- Speech signal: wav file

- Transcription: txt or TextGrid

```
assis sur le mur du jardin potager
j' ai mesuré l' ampleur des dégâts
les choux avaient été entièrement dévorés par les limaces
le potager était complètement dévasté
et ressemblait à un terrain en friche
mais pourquoi est-ce_que j' ai pas pensé à mettre du tue limaces
au point où j' en suis si je m' écoutais je ferais tout cimenter
comme ça j' aurais une belle cour intérieure et plus de soucis
```
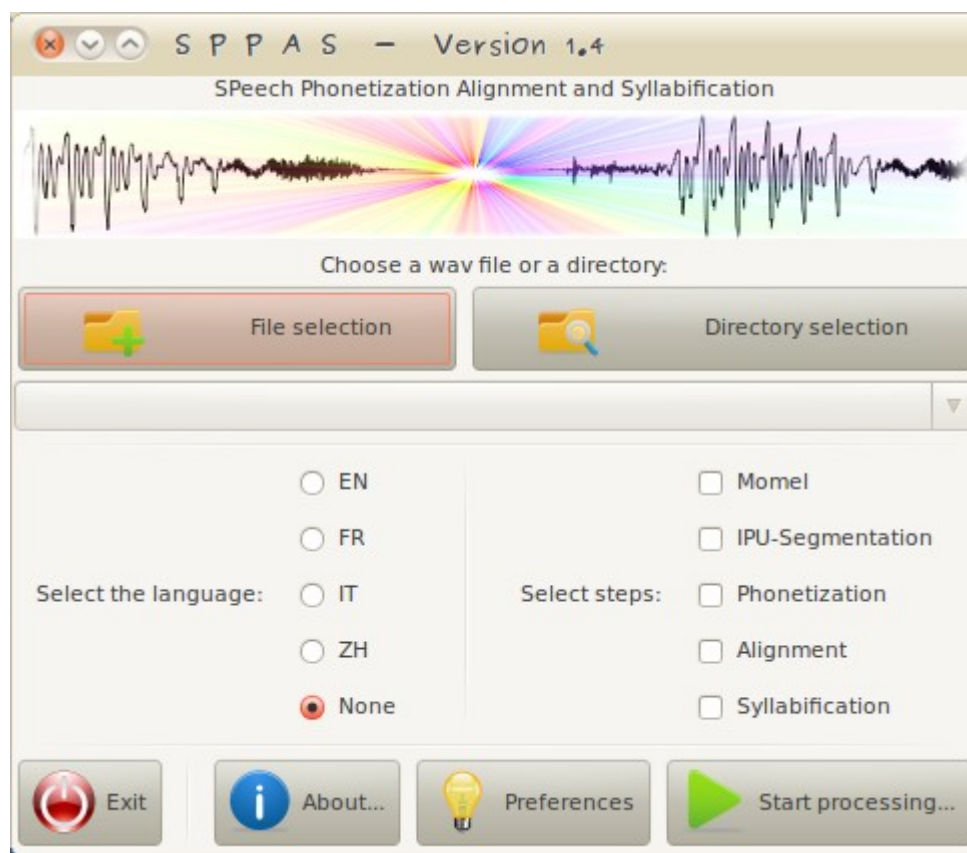
SPPAS outputs

- A set of TextGrid files
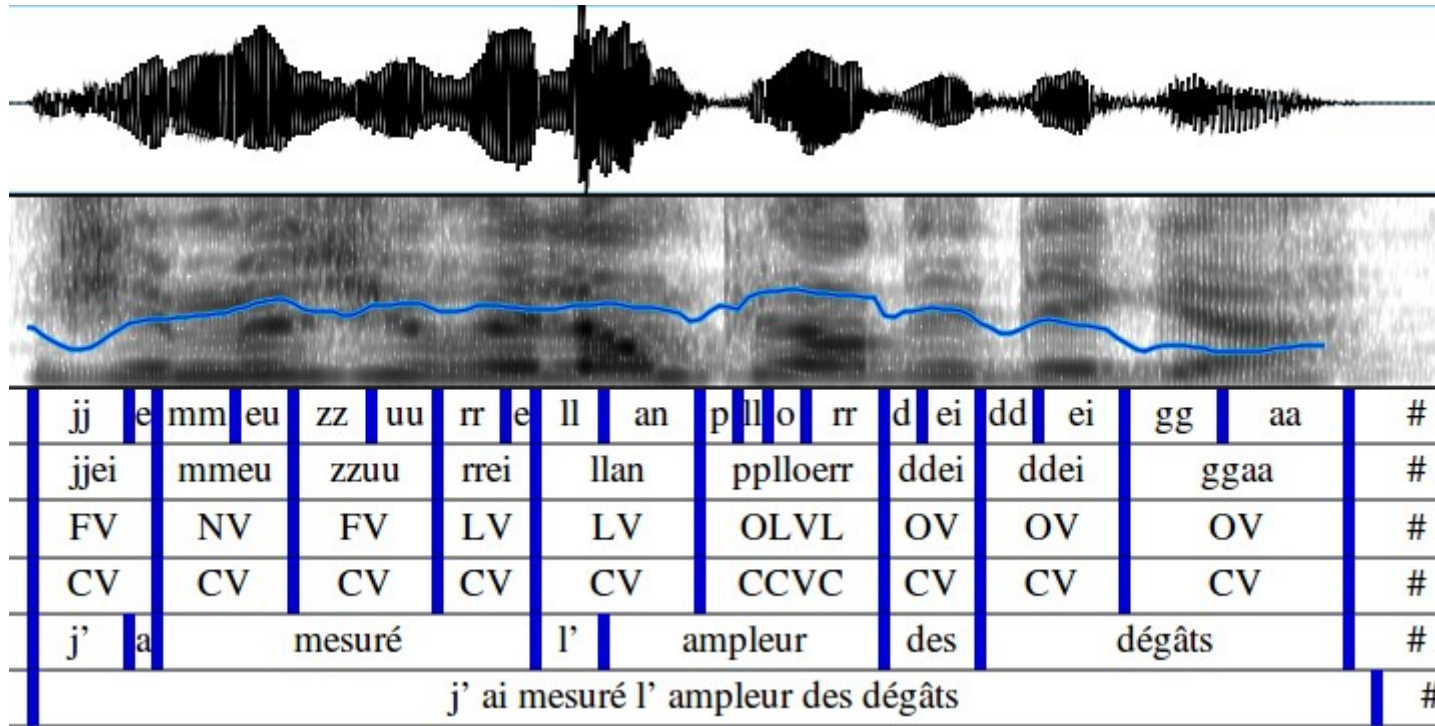
Watch it!

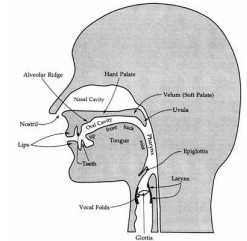# SPPAS in the *Speech tools* community

Computer scientists

Phoneticians Prosodists

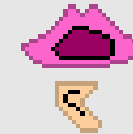| Doing speech technologies with computers	Doing phonetic/prosody with computers
Public tools → **Julius CSR** Local tools Mixed Public/Local tools	Public tools → **Praat** Other tools
Always automatic annotations	Most of time manual annotations

SPPAS

IPU segmentation

- Inter-Pausal Units segmentation

- The algorithm computes a heuristics based on the dectection of silences
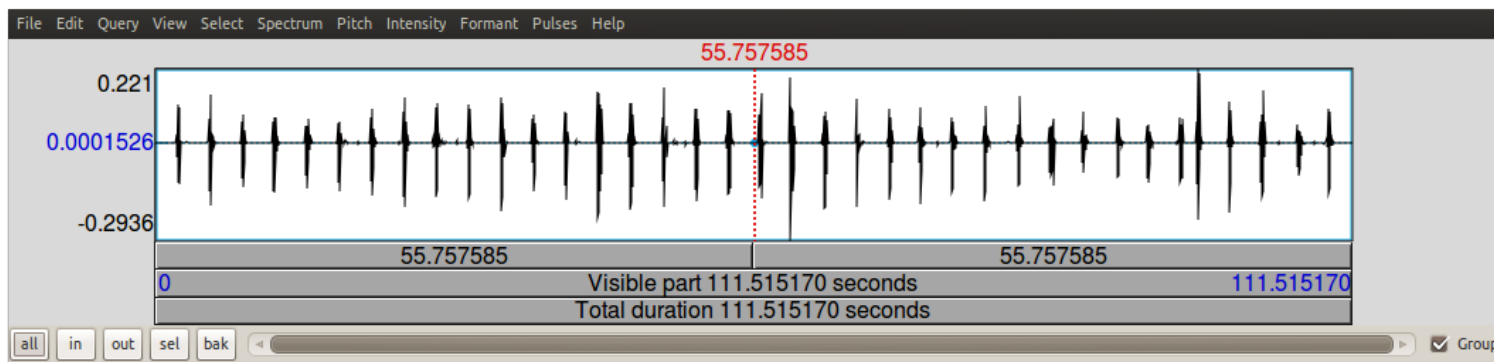
Phonetization

- Phonetization is the process of representing sounds with phonetic signs

- There are two general ways to construct a phonetization process:

  - rule based systems (with rules based on inference approaches or proposed by expert linguists);

  - dictionary based solutions which consist in storing a maximum of phonological knowledge in a lexicon.

SPPAS Phonetization

- SPPAS uses the dictionary-based approach

- The phonetization is the equivalent of a sequence of dictionary-look-ups:

    - Input transcription needs to be word-segmented

    - It is supposed that all words of the transcription are mentioned in the pronunciation dictionary:

        - SPPAS 1.3: Unknown words are labelled: "UNK" and the utterance is not aligned

        - SPPAS 1.4: Unknown words are phonetized automatically

Phonetization: variants

- No rules are applied, all possibilities are stored

    je : `jj|jj.eu|ch`

    suis : `ss.yy.ii|ss.yy.ii.zz|ss.uu.ii|yy.ii|yy.ii.zz`
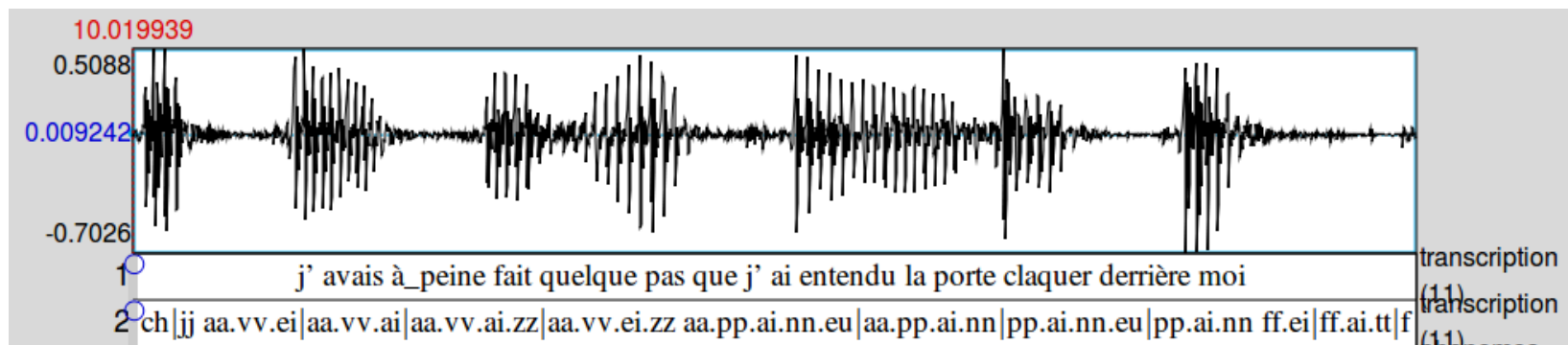
Alignment

- A time-matching between a given speech utterance along with a phonetic representation of the utterance
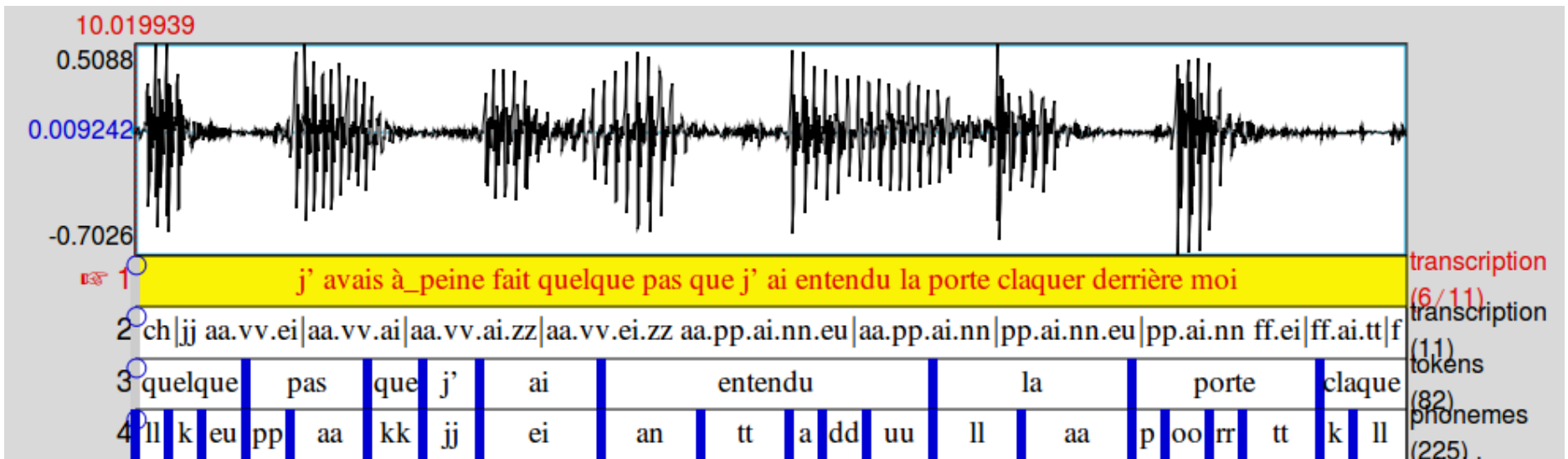
SPPAS alignment

- Forced-alignment in SPPAS is based on the Julius Speech Recognition Engine:

    - A finite state grammar that describes sentence patterns to be recognized;

    - An acoustic model.

- The alignment task is a 2-step process:

    - the first one: choose the phonetization;

    - the second one: perform the segmentation.

Syllabification

- Development of a Rule-Based System for automatic syllabification of phonemes' strings

- The syllabification is based on 2 principles:

  - a syllable contains a vowel, and only one;

  - a pause is a syllable boundary.

- These two principles bring up the problem of finding the boundaries between two vowels.

$$V \, C \, C \, V$$

SPPAS-French syllabification

- Group phonemes into 6 classes: Vowels, Glides, Liquids, Occlusives, Fricatives, Nasals

- Evaluation: the test corpus is 1.6% of CID

    - about 7 minutes of a dialogue;

    - 2068 syllables.

- Syllable agreement rate:

    - 97.8% between 2 experts

    - 95.8% and 94.9% between automatic system and each expert

B. Bigi, C. Meunier, I. Nesterenko, R. Bertrand
Automatic detection of syllable boundaries in spontaneous speech
Language Resource and Evaluation Conference, pp 3285-3292, La Valetta, Malte, 2010

SPPAS: towards prosody annotation

- Momel and Intsint integrated in SPPAS 1.4
  - momel: modélisation de la mélodie
  - Intsint1 and Intsint2 (ask Daniel for details)
- But... today's version:
  - a tool to calculate pitch is missing!

SPPAS resources

- Dictionary:

```
264377 PESAVO [PESAVO] p e s a v o
264378 PESCA [PESCA] p E s k a
264379 PESCA(2) [PESCA] p e s k a
264380 PESCADOR [PESCADOR] p e s k a d o r
264381 PESCAGGI [PESCAGGI] p e s k a dZ i
264382 PESCAGGIO [PESCAGGIO] p e s k a dZ o
264383 PESCAI [PESCAI] p e s k a i
264384 PESCAIA [PESCAIA] p e s k a ja
```

- Acoustic model:
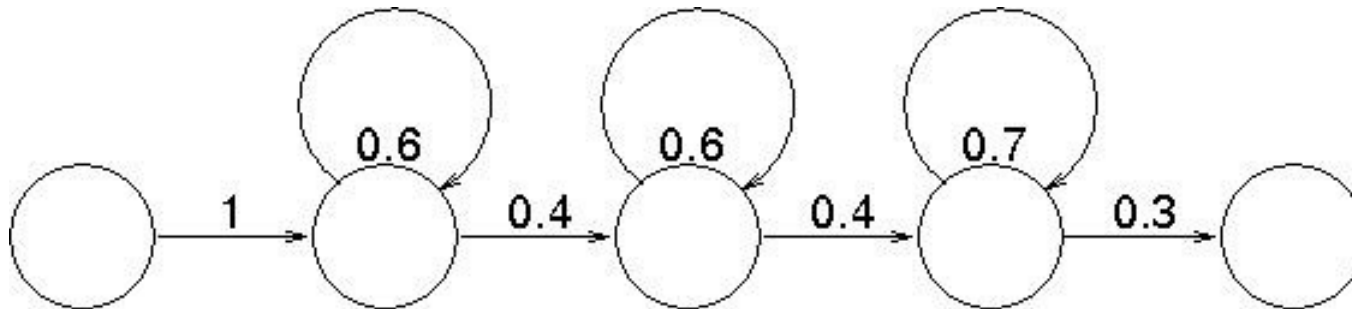
Resources description

	FR	IT	ZH	EN
Dictionary : Number of entries	350k words and 300k variants	390k words and 5k variants	350 syllables	121k words and 10k variants
Acoustic model: Data to train	Triphones - 7h30 CID +30min read	Triphones - 3h30 map-task	Monophones - 50min read	Triphones See voxforge.org

SPPAS in research projects...

- Amennpro:

    - read speech

    - F_F and E_F

- Evalita 2011:

    - Italian phonetization and alignment

- Orthographic Transcription study:

    - which Enrichment is required for phonetization?

Amennpro

- A new corpus named AixOx

  - four way recordings of French and English texts read by English and French speakers

  - each speaker read 40 continuous passages from the *Eurom1* corpus

  - non-native speakers were divided into

    - advanced and beginners

- Alignments for French are in progress

S. Herment, A. Loukina, A. Tortel, D. Hirst, B. Bigi
A multi-layered learners corpus: automatic annotation
4th International Conference on Corpus Linguistics, Jaèn (Spain), March 2012.

Forced-alignment task

EVALITA 2011
Evaluation of NLP and Speech Tools for Italian

EVALITA 2011 is the third evaluation campaign of Natural Language Processing and Speech tools for Italian, supported by the NLP working group of AI*IA (*Italian Association for Artificial Intelligence*) and AISV (*Italian Association of Speech Science*)

http://www.evalita.it/2011

Evalita 2011

- "Forced Alignment on Spontaneous Speech"':

  - Phone segmentation;

  - Word segmentation.

- Dialogues, map-tasks:

  - 3h30 speech;

  - 15% phones are: "sil", filled-pauses, garbage.

- Play examples!

Official Results

- Estimated using sclite, compared to a reference (obtained automatically!):

  - 88.4% good phoneme alignments:

    – This score contains both phonetizations and alignments errors.

  - 96.7% good word alignments.

B. Bigi
Forced Alignment on Spontaneous Speech for Italian: the SPPAS tool
Lecture Notes in Artifical Intelligence, 2012. To be published.

Orthographic Transcription
for phonetization

- Hypothesis:

  - The better transcription is:

    - the better phonetization...
    - thus, the better alignment
    - thus, the better syllabification!

- But... what is a « better » transcription?

- Experiments on French only:

  - Dictionary-based approach (SPPAS)

  - Rule-based approach (the plugin: ESPPAS)

Enriched Transcription: example 1

- Standard Orthographic transcription

- Enriched orthographic transcription:

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

- Very rich orthographic transcription:

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

Enriched Transcription: example 2

- Enriched orthographic transcription:

    - ouais tu comprends na na na na na na la solidarité les étudiants et quelle solidarité ah c'est bon je lui dis tu es solidaire toi tu es solidaire de de de tes fesses tu es solidaire

- Very rich orthographic transcription:

    - ouais tu comprends na na na na na na la solidarité les étudiants et [quelle, qué] solidarité ah c'est bon j(e) [lui,i] dis [tu, ty] es solidaire toi t'es [solidaire,solidaireu] [de,deu] [de,deu] [de,deu] tes [fesses,fèsseu] t'es solidaire

Test corpus, manually aligned

	CID	AixOx	Grenelle
Duration	143s	137s	134s
Nb speakers	12	4	1
Nb Phonemes	1876	1744	1781
Nb Tokens	1269	1059	550
Silent Pauses	10	23	28
Hesitations	21	0	5
Noise, breath...	0	8	0
Laughts	4	0	0
Truncations	6	2	1
Elisions	60	21	43
Special pron.	58	37	23

Phonetization: results

French only system

Optimal

Room for improvements

Language independent algorithms
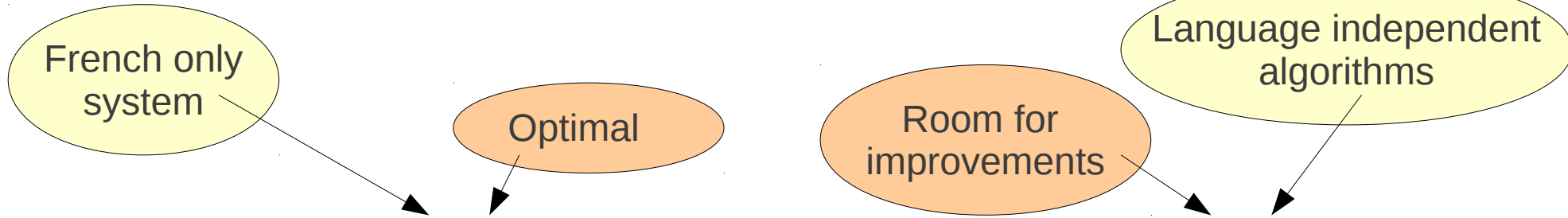
|	ESPPAS:				SPPAS:
	Sub	Del	Ins	Err	Err
CID					
Standard TO	2.8	4.5	10.0	17.3	
Enriched TO	2.7	1.4	10.3	14.4	12.5
Very rich TO	1.8	1.3	3.4	6.5	
AixOx					
Standard TO	1.4	5.0	3.0	9.5	
Enriched TO	1.4	2.3	2.9	6.5	8.2
Very rich TO	1.3	1.8	2.5	5.6	
Grenelle					
Standard TO	1.1	2.8	4.1	8.0	
Enriched TO	1.0	1.2	4.1	6.3	7.2
Very rich TO	1.3	1.0	1.7	4.0	

A few words about technical stuff...

- SPPAS 1.4 accepts only wav audio files,

    - with only ONE channel (mono)

- The transcription encoding must correspond to that of SPPAS dictionary:

    - iso8859-1 for French or Italian,

    - us-ascii for English or Chinese.

- The transcription and the audio files must have the same name (except for the extension)

- One TextGrid = one tier

Perspectives
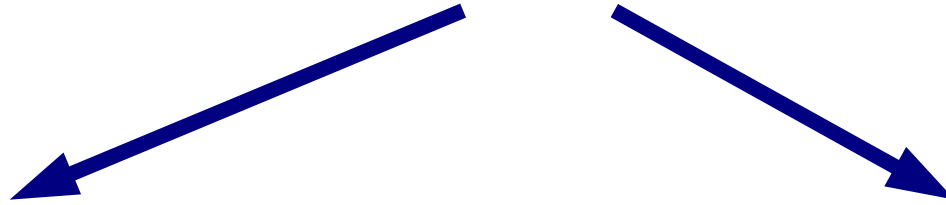
Technical:

Facilitate installation
GUI development
Portability
Forge deposit
...

Research:

Momel in SPPAS
Add new languages
L2
Dialects
Prosody(?) annotation
...

References

B. Bigi, C. Meunier, I. Nesterenko, R. Bertrand. *Automatic detection of syllable boundaries in spontaneous speech.* Language Resource and Evaluation Conference (LREC), pp 3285-3292, La Valetta, Malte, 2010.

B. Bigi. *The SPPAS participation to Evalita 2011.* Working Notes of EVALITA 2011, Rome, Italy, ISSN: 2240-5186, January 2012.

S. Herment, A. Loukina, A. Tortel, D. Hirst, B. Bigi. *A multi-layered learners corpus: automatic annotation.* 4th International Conference on Corpus Linguistics, Jaèn, Spain, March 2012.

B. Bigi, D. Hirst. *SPeech Phonetization Alignment and Syllabification: a tool for the automatic analysis of speech prosody.* Speech Prosody, Shanghai, China, May 2012, Accepted.

B. Bigi, P. Péri, R. Bertrand. *Orthographic Transcription: which Enrichment is required for phonetization?* Language Resource and Evaluation Conference, Istanbul, Turkey, May 2012. Accepted.

B. Bigi. *SPPAS: a tool for SPeech Phonetization Alignment and Syllabification.* Language Resource and Evaluation Conference (LREC), Istanbul, Turkey, May 2012. Accepted.

B. Bigi. *Forced Alignment on Spontaneous Speech for Italian: the SPPAS tool.* Lecture Notes in Artifical Intelligence, Springer, 2012. To be published.

http://www.lpl-aix.fr/~bigi/sppas/