Expressive Speech CORE

Exceptions lexicon

Reference manual

Core version: 8.1

Reference	:	VOX170_exceptions_lexicon_manual_8.1_1.0_EN			
Doc. version	:	1.0	**Status**	:	Release
Date	:	11/04/2017	**Diffusion**	:	restricted

Review

Version	Date	Author	Verified by	Modification
1.0	11/04/2017	ER	PYJ	Création for Expressive Speech Core 8.1

Summary

1 Presentation

1.1 Overview

An exceptions lexicon file is used to associate new and/or corrective information (orthography, pronunciation, grammatical tag, and lemma) with a token (word) that is processed incorrectly by the text-to-speech engine. This may typically be the case of acronyms.
Voxygen Expressive Speech Core accepts any lexicon that is conformant with the Pronunciation Lexicon Specification (PLS) version 1.0 as described in the W3C Recommendation 14 October 2008: http://www.w3.org/TR/pronunciation-lexicon/. The syntax of a PLS file is outlined and details of Voxygen Expressive Speech Core restrictions/extensions to the W3C recommendation for PLS are given in section II.
A legacy proprietary format (called 'EXC') is also supported by Voxygen Expressive Speech Core and described in section III. This depreciated format may still be used but the richer PLS format is recommended.
The input sentence is split into a number of tokens, whereby a token is either a sequence of white spaces, a sequence of alpha-numeric characters (including alphabetic characters with diacritics), or any other single character (a symbol, a punctuation mark, a control character, etc.). Note that there are thus three types of token possible.
Sequences of input tokens are processed in such a way that matches with an entry in an exceptions dictionary can only be made by a set of complete tokens. Thus, partial tokens do not constitute a potential match; for example, the input "n-gram" does not match the entry "en-gram" (the token 'en' cannot be split). The input "Dr" does not match the entry "Dr." (insufficient tokens), but the input "Dr.!" (three tokens) does match the entry "Dr." (plus residual "!").
However, a token of any white spaces in the input text matches a set of any white spaces in an entry.
Whenever a consecutive sequence of tokens matches a lexical entries grapheme, the entries information is attributed to the tokens.

1.2 Terminology

Abbreviation	Description
SSML	*Speech Synthesis Markup Language*
TTS	*Text To Speech*
Baratinoo	*Abbreviation for the Voxygen Expressive Speech Core (Voxygen TTS engine)*

1.3 Reference documents

Reference	Document Name
V0X31	SSML reference manual
V0X32	BARATINOO tags reference manual
V0X349	Phonemes and visemes reference manual

2 PLS Syntax

2.1 PLS file format

A PLS file must be a conformant XML document. It is recommended to always specify the document's character encoding scheme with the `encoding` attribute in the `xml` prologue.

When a PLS document is not compliant, the Baratinoo engine will issue notification of the problem encountered and may return an input error.

2.2 PLS elements

Here is a list of supported PLS elements and their attributes. See the W3C PLS 1.0 recommendation for more details.

The second column ("St") of the attribute tables indicates the status of each attribute:

`M`: mandatory (must/required; error if not present)
`R`: recommended (should; warning if not present)
`O`: optional (may; ok if not present)
`E`: Baratinoo extension (optional)

Baratinoo provides some proprietary extensions to the PLS recommendation:

- Elements: `<lemma>` and `<audio>`
- Attributes: `opt` for `<lexicon>`; `opt`, `say-as` and `scope` for `<lexeme>`; `xml:lang` for `<alias>`

A specific namespace has been defined for extension elements and attributes:

http://www.voxygen.fr/tts

This document suggests using `vox` as the namespace prefix. Please refer to the XML namespace recommendation for details: http://www.w3.org/TR/REC-xml-names/.

Any XML elements or attributes that are encountered from a declared namespace other than those defined within PLS and our extension namespaces will be ignored.

The PLS markup consists of the following elements and attributes:

Elements	Attributes	Description
<lexicon>	version xml:base xmlns xml:lang alphabet *vox:opt*	root element for PLS
<meta>	name http-equiv content	Element containing meta data
<metadata>		Element containing meta data
<lexeme>	xml:id role *vox:opt* *vox:say-as* *vox:scope*	the container element for a single lexical entry

Elements	Attributes	Description
<grapheme>		Contains orthographic information for a lexeme
<phoneme>	prefer alphabet	Contains pronunciation information for a lexeme
<alias>	prefer *xml:lang*	Contains acronym expansions and orthographic substitutions
<example>		Contains an example of the usage for a lexeme
<*vox:lemma*>		Contains lemma information for a lexeme
<*vox:audio*>	src prefer gain speed tempo	empty element providing audio source for a lexeme

<lexicon>

Description

Root of a PLS document.

Attribute	St	Value
`version`	M	`"1.0"`
`xml:base`	O	Specify the base URI of the root document.
`xmlns`	M	`"http://www.w3.org/2005/01/pronunciation-lexicon"`
`xmlns:xsi`	R	`"http://www.w3.org/2001/XMLSchema-instance"` This is required if xsi:schemaLocation is to be given.
`xsi:schemaLocation`	R	`"http://www.w3.org/2005/01/pronunciation-lexicon http://www.w3.org/TR/pronunciation-lexicon/pls.xsd"`
`xmlns:vox`	O	`"http://www.voxygen.fr/tts"` This is required if Baratinoo extensions are used.
`alphabet`	M	Pronunciation alphabet used in the document. Supported values are: `"x-voxygen"` or `"ipa"`
`xml:lang`	M	IETF BCP47 language tag: http://www.ietf.org/rfc/bcp/bcp47.txt
`vox:opt`	E	Set default token matching behaviour. `"i"` or `"I"` make token matching case insensitive. `"!i"` or `"!I"` make token matching case sensitive. `"d"` or `"D"` token matching is to ignore differences in diacritics. `"!d"` or `"!D"` token matching is not to ignore differences in diacritics. The default value is `"!i!d"`

Possible content
Zero or more <meta> elements, followed by an optional <metadata> element, followed by zero or more <lexeme> elements

<meta>

Description

Empty container by which information about the document can be provided.

Attribute	St	Value
http-equiv name	M	A property name
content	M	Value to be associated with the named property

Possible content
Nothing

Restrictions

This element is ignored.

<metadata>

Description

Container in which information about the document can be placed, using a metadata schema.

Possible content
XML data (the content is ignored)

Restrictions

This element is ignored.

<lexeme>

Description

Container for a lexical entry.

Attribute	St	Value
xml:id	O	A unique identifier for the current element. Not used by Baratinoo.
role	O	Used to select the relevant pronunciation of a word when multiple occurrences are available. Works with the SSML <token> element (see below)
vox:opt	E	Override the default token matching behaviour. "i" or "I" make token matching case insensitive. "!i" or "!I" make token matching case sensitive. "d" or "D" token matching is to ignore differences in diacritics. "!d" or "!D" token matching is not to ignore differences in diacritics.
vox:say-as	E	Works with the SSML <say-as> element (see below).
vox:scope	E	"global" (default) "internal" "external"

Possible content
One or more <grapheme> elements, one or more pronunciations (by <phoneme>, <alias> or <vox:audio> elements or a combination of them), zero or more <example> elements and zero or more <vox:lemma> elements.

Note

The children of the <lexeme> element may appear in any order, but note that the order will have an impact on the treatment of multiple pronunciations (see PLS Section 4.9).

The <lexeme> element may contain more than one <grapheme> element to define the base orthography and any variants. Note that all the pronunciations given within the <lexeme> apply to each and every <grapheme>.

The role attribute

When input tokens are explicitly marked by a SSML <token> element, an initial search is done to match exactly the value of the <lexeme> element's role attribute with that of the <token> element's role attribute. If unsuccessful, then a second search is done ignoring role.

Values of the role attribute that are in the vox namespace may inform Baratinoo how to interpret the lexeme; these may be grammatical tags.

The vox:say-as attribute

When the <say-as> element is used in SSML markup, an initial search is performed in a PLS document for lexemes for which the value of the <lexeme> vox:say-as attribute matches the value of the <say-as> element's interpret-as attribute. If unsuccessful,

then a second search is done, ignoring the value of the `interpret-as` attribute; lexemes with no `vox:say-as` attribute may match the `<say-as>` element content but lexemes with a (different) value in the `vox:say-as` attribute will not match.

The vox:scope attribute

The pronunciations of a lexeme with `internal` or `global` scope may be used to determine the pronunciation of the content of `<alias>` elements. The information contained in a lexeme with `external` or `global` scope may be used when matching input tokens.

<grapheme>

Description

The `<grapheme>` element contains text describing the orthography of the lexeme.

The `<grapheme>` element has no attributes.

Possible content
Only text (must not be empty)

<phoneme>

Description

The `<phoneme>` element contains text describing how the `<grapheme>` is pronounced.

Attribute	St	Value
`alphabet`	0	Pronunciation alphabet used for this phoneme element only. Supported values are: `"x-voxygen"` or `"ipa"`
`prefer`	0	`"true"` or `"false"` Indicates the preferred pronunciation to be used by a speech synthesis engine when it is set to `"true"`

Possible content
Only “characters” from the pronunciation alphabet valid for the language (must not be empty)

Note

See VOX349 document for specifications of x-voxygen pronunciation alphabets. Syntax of x-voxygen phonemic strings differs between supported languages:

English

Each phoneme in the string must be separated by a blank.

Syllable, morph and word boundaries can be set in the phonetic content using respectively the characters: '`.`', '+' and '`|`'. Lexical stress can also be specified, using the characters: '*', '~' and '-', respectively for primary, secondary and tertiary accent.

Other languages (French, Spanish, Italian, German, Arabic, etc...)

Phonemes are joined. If a phoneme has two characters long, it must be followed by a " character, e.g. phonemes /F,R,AI,N,CH/ should be written FRAI"NCH".

The sequence ## can be use in the <phoneme> content to mark a word boundary, e.g.

```
voice communication : [VAU"IS##KOMYUNIKAI"CH"YAU"N]
```

Break

A break (short pause) can be inserted between two words with the '%' symbol . The duration of the pause may be lengthen by concatenating '%' symbols; e.g.

```
<lexeme>
      <grapheme>This is a break</grapheme>
      <phoneme> dh"is##iz##%@%%##brei"k</phoneme>
</lexeme>
```

Allow silences to be specified at word-boundaries,

(.) Short pause

(..) Medium pause

(...) Long pause

(x.y) x.y-second pause [max 60.0]

Baratinoo extension

Baratinoo introduces a new attribute to control the inclusion or exclusion of specific acoustic units as candidate realisations for each part of the given phonetic pronunciation.

Attribute	St	Value
vox:idl	E	[*ids*]*pho*[*ids*]*pho*…[*ids*] *ids* = list of comma separated acoustic unit identifiers (integers) preceded by + for 'inclusion', otherwise 'exclusion' is inferred. *Pho* = a x-voxygen phonetic symbol

<alias>

Description

The <alias> element is used to indicate the pronunciation of an acronym or an abbreviated form, in terms of other orthographies, or other substitutions as necessary.

Attribute	St	Value
prefer	O	"true" or "false" Indicates the preferred pronunciation to be used by a speech synthesis engine when it is set to "true"
xml:lang	E	IETF BCP47 language tag. Specifies the written language of the contained text.

Possible content
Only text (must not be empty)

<example>

Description

Provides an example sentence that illustrates an occurrence of the lexeme.

The <example> element has no attributes.

Possible content
Only text (must not be empty)

Restrictions

This element is ignored.

<vox:lemma>

Description

Specifies the lemma of the current lexeme.

The <vox:lemma> element has no attributes.

Possible content
Only text (must not be empty)

<vox:audio>

Description

Specifies the audio source that can be used for the pronunciation of the current lexeme.

If the audio source given by the <vox:audio> element cannot be rendered (not located or opened) then an alternative pronunciation will be used if present, with priority being given to the alternatives by document order.

Attribute	St	Value
src	M	Name of file (absolute or relative URI)
prefer	O	"true" or "false" Indicates the preferred pronunciation to be used by a speech synthesis engine when it is set to "true"
gain	O	Signed number followed by "dB" for decibels. Default "+0.0dB".
speed	O	Unsigned positive number or zero followed by "%". Default "100%".
tempo	O	Unsigned positive number or zero followed by "%". Default "100%".

Possible content
Must be empty.

Restrictions

On some platforms, an HTTP/FTP client is embedded in Baratinoo to download audio files

from a server. URI schemes other than `file:`, `http:` and `ftp:` in the value of the `src` attribute (after application of xml:base) are not supported.

2.3 Example

Here is the beginning of an exceptions lexicon file, for French voices:

```
<?xml version="1.0" encoding="utf-8"?>
<lexicon version="1.0"
     xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-
lexicon
     http://www.w3.org/TR/pronunciation-lexicon/pls.xsd"
     xmlns:vox="http://www.voxygen.fr/tts"
     alphabet="x-voxygen"
     xml:lang="fr">


  <lexeme vox:opt="i">
    <grapheme>n°</grapheme>
    <phoneme>NUMEI"RAU"</phoneme>
    <vox:lemma>numéro</vox:lemma>
  </lexeme>
  <lexeme role="vox:NCoesm">
    <grapheme>2D/3D</grapheme>
    <phoneme>DEU"DEI"TRWADEI"</phoneme>
  </lexeme>
  <lexeme role="vox:NCoe.m">
    <grapheme>A/R</grapheme>
    <phoneme>AAI"R</phoneme>
    <alias>aller-retour</alias>
  </lexeme>
</lexicon>
```

3 Proprietary syntax

3.1 Overview

An EXC exceptions lexicon is a simple text file organized by line.

The file contains an encoding line followed by other lines that may be a comment or a token exception. Empty (white space only) lines are ignored.

During the lexical search, precedence is given by document order; i.e. an exception specified at the beginning of the lexicon has precedence over exceptions specified further on.

3.2 Encoding line

This line must always be present and must be the first uncommented line of the file.

It specifies the encoding character scheme used within the file.

Encoding utf8 can be used for all language.

Encoding `cp1252` (MS codepage close to ISO latin-1) is recommended for French, Spanish and English voices. Encoding `cp1256` is recommended for Arabic. Other supported encoding schemes are: `iso-latin-1, iso-latin-2, iso-latin-6, iso-latin-15, iso-latin-16, cp437 and cp840`.

3.3 Comment line

A comment line always starts with the two characters: `//`

3.4 Token exception line

The syntax of a token exception line is:

grapheme `:` *output* `[`*role*`]` `[`*options*`]` `[`*// comment*`]`

`[]` indicates optional arguments.

- *grapheme* is the exception token(s); e.g. `TTS` or `voice communication`. If there is the character '`:`' or '`\`' in the input, it must be escaped with the character '`\`', e.g. `http\:/www`
- *output* is either a new orthography (alias) or a pronunciation (phoneme).
  - A new orthography must be enclosed by `< >`; e.g. `<Text to speech>`

    Pronunciation of the new orthography is *not* determined from the content of other exceptions specified in the file.
  - A pronunciation be enclosed by `[ ]`; e.g. `[FRAI"NCH"]`

    Pronunciations are given using the x-voxygen pronunciation alphabets.

    Sequence ## can be use in a pronunciation to mark a word boundary; e.g.

    `voice communication : [VAU"IS##KOMYUNIKAI"CH"YAU"N]`
- *role* is an optional grammatical tag enclosed by `( )`; e.g. `(NCoesm)`

- *options* is a list of options that control lookup. Possible values are:
  - `/i` to ignore case in the input.
  - `/d` to ignore diacritics
  - `/s say-as` to indicate a say-as mode required for the exception.
- `//` *comment* is a comment.

3.5 Example

Here is the beginning of an EXC exceptions file, for a French voice:

```
cp1252

n° : [NUMEI"RAU"] /i
ABC : [ABEI"SEI"](NCoesm)
API : [APEI"I] /s sigle
voice communication : [VAU"IS##KOMYUNIKAI"CH"YAU"N] /i
```