

Position paper for the Second W3C Workshop on Internationalizing the Speech Synthesis Markup Language (SSML)

Subject: Extensions to the <say-as> and <sub> tags for Inflected Languages
Date: 2006.04.13
Authors: Chris Vosnidis (vosnidis@speech.gr)

1. Motivation

Greek is a heavily inflected language. Articles, nouns and adjectives usually find themselves in agreement with regard to their number, case and gender within the context of a phrase. Numbers, both cardinals and ordinals, and abbreviations are the two most prominent categories of non-normalized text entities whose expansions should be determined by their surrounding context. When found either on their own or combined together, as for instance in the case of measurements, they should be expanded to a form that agrees in number, case and gender with their context.

Certain synthesis processors have implemented ad-hoc methods for determining the inflection attributes that should be applied in such text-normalization cases. Such methods usually make certain assumptions with regard to the syntax, and try to locate the group of words determining the inflection attributes that will characterize the normalized output. However, given the fact that the Greek language is also quite free in its syntax, these approaches are very limited in their applications. We believe that SSML should provide a method for explicitly applying such attributes on the non-normalized text, allowing the synthesis processor to determine the expansion that should be used.

2. Proposed extensions for the <say-as> tag

2.1. The inflection attribute

We propose the introduction of the 'inflection' attribute to the <say-as> tag. The attribute contains a series of inflection specifiers, which can be used by the synthesis processor as hints determining the number, case and gender of the <say-as> tag content. The attribute is valid for most of the types of <say-as> content. It will however be most helpful for number-related content.

The attribute's syntax is described below:

```
inflection=" ?<case> ?<gender> ?<number>"
```

where:

```
<case>:  ['nominative', 'accusative', 'dative', 'genitive',  
         'vocative']  
<gender>: ['masculine', 'feminine', 'neuter']  
<number>: ['plural', 'singular']
```

Attribute content element order is irrelevant of the attribute's interpretation. Not all content elements are required.

The following code provides an example of the attribute's usage.

```
Input:
```

```
<say-as interpret-as="cardinal" inflection="nominative feminine">
3</say-as> πατάτες
```

Output:

τρεις πατάτες (three potatoes)

BUT

Input:

```
<say-as interpret-as="cardinal" inflection="nominative neuter">
3</say-as> παιδιά
```

Output:

τρία παιδιά (three children)

The inflection attribute is even more useful in the case of ordinals, which are fully inflected. Actually, this attribute is quite useful for any other inflected language. In German for instance:

Input:

```
<say-as interpret-as="cardinal" inflection="feminine nominative">
3</say-as> Stunden
```

Output:

drei Stunden (three hours)

BUT

Input:

```
<say-as interpret-as="ordinal" inflection="feminine nominative">
3</say-as> Stunde
```

Output:

dritte Stunde (third hour)

2.2. Proposed legal values for the details and format attributes.

2.2.1 Use of *format* in <say-as interpret-as="telephone">

Digits in phone numbers are grouped together in clusters in order to be announced. In some languages these clusters are read out as digit sequences, whereas in others they are pronounced as numbers. Although grouping can be achieved by utilizing special formatting characters ('-', '.', ' '), such manipulation might prove difficult, especially in cases where the numbers are extracted from a data source.

We propose the utilization of the say-as format attribute in order to define such grouping information. The current SSML standard ignores this attribute within the context of "interpret-as='telephone'". The attribute's proposed content is a series of digits. Each digit determines the length of a digit group, and the sum of all these digits should be equal to the length of the say-as element's text content. The syntax is described below:

```
format="n+"
```

The following code provides an example of the attribute's usage:

```
Input:
<say-as interpret-as="telephone" format="3223" details="30">
  2105552121
</say-as>
```

```
Output:
four digit groups: '210', '55', '52', '121'
```

The digit groups may be interpreted as numbers or digit sequences. This is a language specific decision, and the synthesis processor is responsible for implementing the appropriate behavior.

2.2.2 Use of *details* in `<say-as interpret-as="date">`

The 'date' value of the interpret-as attribute is quite useful in disambiguating the input as a date, with the format attribute providing hints with regard to the way that the input string should be parsed. However, there is no method that allows the user to specify the manner in which the date will be rendered. The current SSML standard ignores the "details" attribute within the context of "interpret-as='date'". We propose the use of the details attribute in order to pass this information onto the synthesis processor. The proposed syntax for the details attribute is the following:

```
details="??<day>??<month>??<year>"
```

Any of the three detail parts are optional, and their position interchangeable.

Some sample values:

```
<day>
- dddd: day of the week
- d: day of month
<month>
- m: month (number)
- MMMM: month name
<year>
- y: year (short form: '98)
- yyyy: year (full form: 1998)
```

Such information will be useful in order to announce only parts of the date, such as the month or the year, or combinations of those, or even change the method by which the month is announced. The following examples are indicative of the proposed use:

```
<say-as interpret-as="date" format="ymd" details="dMMMMyyyy">
  2006-04-03
</say-as>
```

```
Output:
τρεις απριλίου δύο χιλιάδες έξι (Greek)
dritten april zwei tausend sechs (German)
```

```
<say-as interpret-as="date" format="ymd" details="dmyyyy">
  2006-04-03
</say-as>
```

```

Output:
τρεις τειάρτου δύο χιλιάδες έξι      (Greek)

<say-as interpret-as="date" format="ymd" details="MMMMyyyy">
  2006-04-03
</say-as>

Output:
απρίλιος δύο χιλιάδες έξι          (Greek)

```

3. Proposed tag: <aliasmap>

Abbreviations usually stand for nouns and adjectives, i.e. inflected words. This means that the synthesis processor should be able to normalize them using the appropriate case, gender and number information. The current standard provides support for the <sub> element, allowing the replacement of all instances of a token by a single expansion, regardless of their context.

We propose the introduction of an optional "aliasmap" attribute to the <sub> element, in order to link a token to a series of alternative expansions. We also propose the introduction of the <aliasmap> element, which will provide a series of expansions of a given token, distinguished by the contents of the inflection attribute. Token substitution is performed by looking up the aliasmap contents for the given token, and determining the expansion whose inflection attribute matches more closely the inflection attributes of the context.

The proposed syntax is the following:

```

<sub aliasmap="token">
  token
</sub>

<aliasmap id="token">
  <alias inflection="<case> ?<gender> ?<number>?">
    expansion1
  </alias>
  <alias inflection="<case> ?<gender> ?<number>?">
    expansion2
  </alias>
</aliasmap>

```

The proposed nomenclature is adapted towards W3C's Pronunciation Lexicon Specification recommendation.

The proposed extension is useful for processing non-normalized items such as measurements. For instance:

```

<sub aliasmap="κ.">
  κ.
</sub>

<aliasmap id="κ.">
  <alias inflection="plural nominative">
    κιλά

```

```
</alias>
<alias inflection="plural genitive">
  κιλών
</alias>
</aliasmap>

Input:

<say-as interpret-as="cardinal" inflection="plural neuter genitive">
3 κ.</say-as>

Output:
τριών κιλών (three kilos)
```

In this example, the requested inflection context contained the values "plural", "neuter" and "genitive". Since aliasmap's contents do not provide a gender modifier, the "neuter" value is ignored, and the closest match is the entry with inflection values "plural genitive".

3. Company profile

Dialogos Speech Communications S.A. (www.speech.gr) is a pioneering speech technology company located in Chania, Greece. Dialogos was founded in 1998. Its activities focus on the development of automated speech-enabled systems that provide information and services over the telephone and the World Wide Web. Dialogos also conducts basic and applied research for the enhancement of speech recognition and speech synthesis (TTS) systems. Dialogos employs linguists and speech recognition and synthesis experts, and has partnered with Nuance in order to adapt its ASR system to Greek, Turkish and Arabic and its TTS system to the Greek and German languages.

References

- SSML 1.0 say-as attribute values - W3C Working Group Note 26 May 2005 (<http://www.w3.org/TR/2005/NOTE-ssml-sayas-20050526>)
- Pronunciation Lexicon Specification (PLS) Version 1.0 - W3C Working Draft 31 January 2006 (<http://www.w3.org/TR/2006/WD-pronunciation-lexicon-20060131/>)
- Speech Synthesis Markup Language (SSML) Version 1.0 - W3C Recommendation 7 September 2004 (<http://www.w3.org/TR/2004/REC-speech-synthesis-20040907/>)