Integrating corpus-linguistic and conversation-analytic transcription in XML

Abstract

In this short paper I wish to sketch out the research opportunities that come with the recent addition to BNCweb of very large numbers of audio files for the spoken component in the BNC. I argue that the availability of the audio files enables researchers not only to correct the orthographic transcripts, but also to re-transcribe the conversations using conversation-analytic transcription. Finally, I demonstrate that the CA transcripts can be integrated into the BNC’s XML annotation network and illustrate how XML query tools such as XPath and XQuery can be used to efficiently exploit the XML network.

Keywords: BNC, BNCweb, conversation-analytic transcription, XML, XPath, XQuery, corpus pragmatics

1. Introduction

The British National Corpus (BNC) is no doubt a remarkable success story as it, probably, represents the most widely-used corpus in corpus-linguistic research. A large contribution to its success has been the creation of BNCweb, a web interface for the corpus (Hoffmann et al. 2008). BNCweb recently added yet another achievement to its already long list of achievements: a substantial number of the audio recordings from which the transcripts were made have been made available through the interface (Coleman et al. 2012). Researchers can now examine most of the spoken texts, hitherto only accessible through the orthographic transcripts, by listening to them; audio files can be copied and exported into audio processing software for further analysis. This is an eminent new research opportunity. In this short research note the aim is to briefly sketch out this opportunity.

To start with, transcription errors in the text can be corrected thus improving the verbal record. Also, speakers, up until recently hidden behind speaker Id tags, come to life, their voices can be heard and speech delivery can now be examined first-hand: voice quality and its modulation in mimicry, changes in volume, shifts in pitch, slowing down or speeding up — everything is out in the open. Finally, characteristics of sequencing can be determined: pauses can be measured down to split seconds, latching can be detected, overlap can be ascertained beyond doubt.

In other words: there exists now a resource that invites and facilitates the kind of fine-grained transcription that Conversation Analysis has made its hallmark. Considering that the audio files record hundreds of hours of conversations involving very large numbers of very diverse speakers, the potential for conversation-analytic research is immense and probably, since it’s all publicly available, without parallel. What’s more, the audio files come with already-complete orthographic transcripts, so the conversations need not be transcribed from scratch. What remains to be done is re-transcribing the conversations to weed out erroneous transcription and align the transcripts with CA conventions. This requires still a lot of work but far less work than a new transcription. Finally, for both the audio data and their transcripts, BNC’s XML architecture is in place providing meta-data related to speakers (age, sex, class, etc.), speaking turns (delimited by <u> elements), and Part-of-Speech (morpho-syntactic analysis of word-class) for each and every word form. That is, altogether three resources are available: the audio files, the transcripts (initially orthographic and, after re-transcription, in CA format), plus the BNC’s XML scaffolding. If XML is used for the orthographic-transcripts-turned-CA-transcripts it will be possible to integrate CA-style transcription in XML format. Also, data in XML format is searchable and extractable with great efficiency using XML query tools, such as XPath and XQuery (for an introduction for corpus linguists see Author 2015a).
2. Orthographic transcript

In the following I illustrate the procedure of re-transcribing audio-based BNC data as CA transcription and integrating the newly-discovered conversation-analytic details into the BNC’s existing annotational architecture. The text chosen for illustrative purposes is a short storytelling from the BNC file KBD. The telling is part of an extended round of stories (cf. Sacks 1992) thematically related to unlucky fishing experiences. Excerpt (1) is the orthographic transcript downloaded from BNCweb; lines are numbered for ease of reference, the numbers between speaker names and text represent counts of s-units in the file, <-|-> denotes the boundaries of overlapped speech; finally, arrows are used to draw attention to special features of the transcript:

(1)
[BNC: KBD 1790-1801]

Alan 1 1788 Well it's, it's luck innit?
2—> 1789 <-|-> I don't know.
Barry 3—> 1790 <-|-> I don't know <-|-> what's going on.
4—> 1791 I don't <pause> we got about three, three thirty in the morning, both of them
5—> went out to er <pause> canal somewhere up <pause> Dulgate, past Dulgate
6 <pause> we set up and we'd <pause> we'd been fishing for about two and
7 half hours it's aba-- <pause> about six thirty in the morning this old farmer
8 comes up says er <pause> aye, aye lads, he said er <pause> I wouldn't
9 bother it, they drained this area of the canal a few <voice quality:
10 laughing>months ago!
11—> 1792 And we said, oh<end of voice quality>!
Alan 12—> 1793 Yeah. <laugh> <-|->
Barry 13 1794 <-|-> <laugh> <-|-> <pause> <voice quality: laughing>Sat there watching our
14 floats for hours<end of voice quality>!
15 1795 I mean
Alan 16 1796 Yeah.
Barry 17 1797 I mean luckily you, you know, you'd gone on a car, with a car so it's a matter
18—> of throwing everything in the back <-|-> and just going <-|->
Alan 19—> 1798 <-|-> That's it.
Barry 20—> 1799 somewhere else so <pause dur="6">could have sat there all bleeding day!
21 1800 And not have known anything about it.
Alan 22 1801 Aye.

Even a cursory look through the excerpt raises a few questions: in line 2 the second overlap delimiter <-|-> is curiously missing. The same absence can be observed in lines 12 and 19. The absence is explained by the fact that some tags got lost in the course of the BNC’s conversion from SGML to XML. These losses affected mostly instances of overlap and pauses (cf. Hoffmann et al. 2008: 57). Given the faulty annotation, automatic retrieval of these instances of overlap is impossible. Further, the transcript records nine pauses; only for one of them, in line 20, the duration is given (6 seconds).
3. CA transcript

The following is an audio-based CA transcript of the same storytelling passage using Jeffersonian transcription symbols (for CA transcription conventions see, for example, Liddicoat 2007, Schegloff 2000; also, see the list of relevant transcription symbols in the Appendix).

(2)
["Drained canal", BNC: KBD 1790-1801]

Alan 1—> Well it's, it's (.) luck innit [( I don' know), ]
Barry 2—> [ I remember ] once go:n' on,
3—> I got- (0.4) we got up 'bou' three three thirty in the morning
4 ( ) went out to er (0.9) canal somewhere up
5 (1.3)
6 Dulga' area past Dulgate
7 (1.3)
8 we set up
9 and we'd we'd been fishing for about two and half hours
10 it's aba- about six thirty in the morning
11 this old farmer comes up
12—> says er (1.1) ↑Aye aye lads,
13 he said er (0.7) I wou' n't bother it
14—> they drained this area of the canal a few months aG(h)O Hhh::::;
15—> [ hh:::: GGAeexh::: he ] he he
Alan 16—> [ huh huh huh huh huh ]
Barry 17—> S(h)at there watching our floats for hours uhheh heh:
Alan 18 yeah
Barry 19—> I mean luckily you- you know, you'd gone on- with a car
20—> so it's a ma'er o' throw'n ev'ryth'n in th' back
21—> [ 'n' j's go:n'] s'm're else sort of (ay) (1.5)
Alan 22—> [ tha's it ]
Barry 23 could've sat there all bleed'n' day!
24 (1.00)
25 and not known anythin' about it.
26—> (4.4)
Alan 27 aye

The changes that have been made to the original transcript are numerous affecting many layers of discourse and interaction. Not all of them can be mentioned in detail here. First, overlap annotation in (1) has been corrected in (2), where it is indicated by square brackets around the overlapped speech; see lines 1-2 and 21-22. Second, laughter, summarily indicated as <laugh> in (1), is fully transcribed in (2) with appropriate vowels and laughter pulses in lines 15 and 16 thus indicating what the laughter 'sounded' like. Third, the verbal record has seen many corrections, some large, some small. For example, Barry's I don't know indicated in (1) in line 3 is not supported by the audio data. Only Alan seems to say I don' know (without producing a hearable t on the negation) in line 1; however, due to its occurring in overlap, it cannot be heard with certainty but represents the transcriber's 'best guess', as indicated by the parentheses. Barry, in the same overlap, says I remember (with stress on l), certainly an important change over I don't know in (1) in that the verb remember is commonly found in story introductions and thus counts among the turn design features projecting a storytelling sequence (Rossano 2013). Also, instead of I don't- in line 4 in (1), we find I got- in line 3 in (2). Perhaps the most important textual correction concerns line eleven in (1): there, Alan appears to use constructed dialog in And we said, oh in line 11. The audio tape does not evidence constructed dialog at this point; the BNC's transcriber may have misheard Alan's laughing delivery of S(h)at in line 17 as said oh. Given the centrality of constructed dialog to storytelling, this correction in (2) must be considered essential.
Fourth, pauses are transcribed rather differently in the two transcripts. While the number of pauses is not dramatically different — there are ten pauses in (2) as opposed to nine in (1) — the durational information is dramatically different: not only are all pauses but the very first one (which is shorter than 0.3 seconds) measured in seconds up to one decimal but also the 6-second pause recorded in transcript (1) has disappeared and given way to a mere 1.5 second pause. What is more, there is a major new pause in line 26 in (2) that was apparently overlooked in (1). This long inter-speaker pause is interactionally significant in that it conveniently signals the storytelling’s completion.

Fifth, transcript (2) is rich in articulatory detail. To begin with, in (1) Alan is transcribed as having used Dulgate twice in its full form, whereas in (2) the first mention of the location in line 3 is shortened to Dulga’. Other such ‘deviant’ pronunciations include don’ in line 2 (instead of don’t) and woul’ in line 13 where the d is silent. In line 12, the upward arrow ↑ indicates a sharp rise in pitch, whereas the upper case A in ↑Aye aye lads marks increased loudness; the colon(s) used in Hhh:::: (line 14) and gan’ (lines 2 and 15) are used to indicate the stretching of the sound preceding them. Italics in I do (line 2) and luckily (line 19) are used to indicate stress. Punctuation, finally, reflects intonational contours, with the period in lines 1, 2, and 25 indicating a falling tone and the comma in lines 12, 14, 16, and 19 indicating incomplete intonation.

In sum, there will be little doubt that transcript (2) is of much better quality due to its greater verbal accuracy and its richer phonological, sequential, temporal and interactional detail. The key question is whether this level of detail defies XML transposition. In other words: can CA transcripts be turned into XML transcripts without compromising dearly held CA principles?

4. XML transcript

In the following I discuss a few key excerpts from the XML version of transcript (2). The annotation scheme is still work-in-progress; critical feedback to the author is very much appreciated. The full current scheme can be found in the Appendix.

The current tagging scheme underlying the XML annotation of audio data provides for five broad categories, in XML parlance ‘elements’: the <seq> element for sequential features including overlap and latching; the <phn> element for phonological characteristics including, for example, volume, stress, pitch, intonation, etc.; the <tmp> element for temporal aspects including pausing and variation in delivery speed; the <laugh> element for laughter including within-speech and between-speech laughter; and, finally, the <cmmt> element for transcriber comments relating, for example, to hearability issues and extra-linguistic events. (For video recordings, more elements are required to capture gaze and gestures.)

In the XML excerpts discussed below, two element types found in the BNC’s XML file are consistently omitted: <s> elements for sentence-like units as well as <c> elements for grammatical punctuation. In the BNC, <w> elements have ‘c5’, ‘hw’ and ‘pos’ attributes. In the interest of legibility of the extracts, the ‘hw’ and the ‘pos’ attributes have been removed. Other omissions (for expository reasons) are indicated by XML comments in the form of <!--  -->. Finally, I also show some rather simple XPath queries for retrieving data from the transcript. The aim here is to illustrate the potential of this technology for extracting data from densely annotated XML documents.

Excerpt (3) represents lines 1-2 of the CA transcript in (2). The excerpt illustrates XML annotation of overlap as a sequence feature and some phonological features:

```
(3)  1 <u who="PS040"/>
  2 <!-- w-elements omitted -->
  3 → <seq type="ovl" n="1" part="fst">
  4 →  <cmmt aud="poss">
  5 →   <phn int="fall">
  6   <w c5="PNP" >I </w>
  7   <w c5="VDB" >do</w>
  8 →   <phn real="n' ">
  9 →     <w c5="XX0" >n't </w>
 10 </phn>
```
In lines 3 and 17, we find <seq> elements (for sequence phenomena) specified by type="ovl" as overlap. The <seq> element is wrapped around the simultaneously produced verbal material. The elements have two more attributes intended to provide handles by which to 'yank out' different types or components of overlap. One is 'n': its values 1, 2, etc. number the overlaps consecutively in their textual environment. With this attribute, overlap elements can be extracted by position in overlap sequences: the first overlap, the second, and so forth. If extracted via 'n', the query will return both the overlapped speech and the overlapping speech. For example,

```
//seq[@n="1"]
```

selects all text-initial overlaps (in the present case, just one).

If, by contrast, the aim is to tease apart overlapping and overlapped speech, the third attribute 'part' comes handy: by specifying part="fst", only overlapped speech will be addressed (in most cases, address forms, laughter, tags, or, as in the present case, turn increments; cf. Jefferson 1979), while specifying part="scd" addresses overlapping speech. For example:

```
//seq[@part="scd"]
```

Within the overlapped sequence, we find in line 4 a <cmmt> element. This element type accommodates the transcriber’s comments related, for example, to extra-linguistic events or audibility issues; in line 4, the attribute aud="poss" indicates that I don’t know is just a candidate hearing. In line 5, we encounter the first <phn> element. <phn> elements capture phonological characteristics. The element in line 5 grasps the intonation contour of I don’t know, which is falling, as indicated by int= "fall". The <phn> element hosts another <phn> element in line 8. This element is specified as real="n' " and it encloses the <w> element for n’t in line 9; the meaning of real="n' " is then that the contraction n’t is deviantly realized as n’. Why make this distinction? If n’ were used instead of n’t in the <w> element, it would appear as a separate entry in a frequency list. This would be counterintuitive because n’ is hardly a different word but rather a different realization of n’t. It should therefore be counted among the occurrences of n’t. This is made possible by its two-fold record both in the <phn> and the <w> element (frequency lists are commonly made from the latter). We notice the same double entry for go:n’ on as a value on the ‘real’ attribute in line 21 and going in line 22. The <phn> element in line 21 also contains the attribute values lth=":o:" and degr=":" indicating the lengthened sound and the degree of lengthening. In line 18, the falling intonation affecting what ‘s go:n’ on is captured using the attribute value int="fall".

We instantly see that the XML annotation has already gained considerably in 'depth': the <u> element in line 1 is parent to the <seq> element in line 3, grandparent to the <cmmt> element in line 4, grandgrandparent to the first <phn> element in line 5, grandgrandgrandparent to three
<w> elements and the second <phn> element and, finally, grandgrandgrandgrandparent to the <w> element which is the child of the second <phn> element. Widely-used corpus tools such as WordSmith fail this depth of annotation; the XPath and XQuery technologies, by contrast, can handle it with ease. For example, if the interest is in deviant pronunciations occurring in the overlaps by a distinct speaker, this rather simple XPath returns, in overlaps by Barry, all <phn> elements that have a ‘real’ attribute wherever they may be tucked in the XML hierarchy:

```
//u[@who="PS03W"]/seq[@type="ovl"]/phn[@real]
```

If researchers wish to retrieve not the elements with all their meta-information but only the sound string, this query using the string() function is hardly more complex:

```
//u[@who="PS03W"]/seq[@type="ovl"]/phn[@real]/string(@real)
```

That CA transcription in XML format need not always cause deep (and challenging) dependencies is shown in extract (4), which represents lines 12 to 13 in the CA transcript in (2). The extract illustrates XML annotation for pausing and variation in delivery:

```
(4) 1  <u who="PS03W">
2    <!-- w-elements omitted -->
3  3 <w c5="VVZ">says</w>
4    2 <w c5="UNC">er</w>
5  3  --> <pause dur="1.1"/>
6  4  --> <phn ptch="up" vol="high" which="A">
7    5 <w c5="ITJ">Aye</w>
8  6  </phn>
9  7 <w c5="ITJ">aye</w>
10 8 <w c5="NN2">lads</w>
11 9  --> <phn int="ctd"/>
12 10 <w c5="PNP">he</w>
13 11 <w c5="VVD">said</w>
14 12 <w c5="UNC">er</w>
15 13  --> <pause dur="1.1"/>
16 14 <w c5="PNP">I</w>
17 15 <phn real="wou' ">
18    16 <w c5="VM0">would</w>
19 17  </phn>
20 18 <w c5="XX0">n't</w>
21 19 <w c5="VVI">bother</w>
22 20 <w c5="PNP">it</w>
23 21  <!-- w-elements omitted -->
24 22 </u>
```

In (4), lines 3 and 13 show elements for pauses; the attribute values dur="1.1" and dur="0.7" record their duration. Note that the pause in line 3 nicely demarcates the beginning of constructed dialog in lines 5-8 thus functioning as auditory quotation marker (cf. Bolden 2004, Author 2013). Moreover, the attribute ptch="up" in line 4 accounts for the sharp pitch rise occurring in Aye, while the attribute values vol="high" and which="A" denote that the beginning of the word was spoken more loudly than the rest of the word. As in this case, sharp pitch rises are presumably always interactionally and discoursally significant. In (4), the significance lies in the initiation of constructed dialog, where the pitch rise, much like the pause, helps the listener identify, to use Goffman's (1981) terminology, the speaker’s switch from ‘author’ (of his own words) to ‘animator’ (of the old farmer’s words). XPath can easily target these pitch rises, simply by addressing the <phn> elements that satisfy the condition of having the attribute value ptch="up":

```
//phn[@ptch="up"]
```
To investigate whether sharp pitch rises collocate with quotative verbs and, hence, with constructed dialog (to my knowledge, an as-yet neglected research question), this query retrieves all words preceding the rise in a 3-word window:

//phn[@ptch="up"]/preceding::w[position()=1 to 3]

Excerpt (5) illustrates how the tagging scheme handles laughter. As is well known, laughter is a central concern in CA research (for example, Jefferson 1985). CA transcription of laughter aims to approximate what the laughter sounds like by paying attention to laughter pulses and the appropriate vowel (Liddicoat 2007: 26). Laughter in conversation can occur in two forms: either as within-speech laughter or as between-speech laughter. The current tagging scheme targets laughter using the <laugh> element. The two types of laughter are distinguished by the attributes 'wsl' and 'bsl' respectively. The laughter in lines 13-14 is within-speech. Two attribute values are used to grasp its character and form: the value desc="aG(h)O" describes the word within which it occurred and vol="high" denotes that it occurred with high volume on the second syllable — a restriction expressed in the which="G(h)O" attribute value. The two remaining laughter occurrences in lines 17-18, 20, and 23-25 represent between-speech laughter most of it produced by Barry and Alan simultaneously. The 'desc' attributes record laughter pulses as well as vowels. The attribute 'dur' records its duration; in lines 17 and 24 the value is dur="2.8" — with almost three seconds a rather long simultaneous laugh. The two laughs' almost complete occurrence in overlap also represents a neat illustration of Schegloff’s analysis of laughter as a ‘choral’ activity “NOT to be done serially (...) but simultaneously” (Schegloff 2000:6).

(5) 1  <u who="PS03W">
  2     <w c5="PNP">they</w>
  3      <w c5="VVD">drained</w>
  4      <w c5="DT0">this</w>
  5      <w c5="NN1">area</w>
  6      <w c5="PRF">of</w>
  7      <w c5="AT0">the</w>
  8      <w c5="NN1">canal</w>
  9      <w c5="AT0">a</w>
 10      <w c5="DT0">few</w>
 11      <w c5="NN2">months</w>
 12 —>  <laugh type="wsl" desc="aG(h)O" vol="high" which="G(h)O">ago</laugh>
 13 —>  <w c5="AV0">ago</w>
 14 —>  </laugh>
 15 </u>
 16     <seq type="ovl" n="2" part="fst">
 17 —>  <laugh type="bsl" dur="2.8" desc="huh huh huh huh huh">huh</laugh>
 18 —>  <w c5="AV0">ago</w>
 19 </seq>
 20 —>  <laugh type="bsl" desc="he he">dur="0.4"/></laugh>
 21 <u>
 22     <w c5="PNP">they</w>
 23      <w c5="VVD">drained</w>
 24      <w c5="DT0">this</w>
 25      <w c5="NN1">area</w>
 26 —>  <laugh type="wsl" desc="aG(h)O" vol="high" which="G(h)O">ago</laugh>
 27 —>  <w c5="AV0">ago</w>
 28 —>  </laugh>
 29     <seq type="ovl" n="2" part="scd">
 30 —>  <laugh type="bsl" dur="2.8" desc="huh huh huh huh huh">huh</laugh>
 31 —>  </seq>
 32 </u>
Finally, the larger storytelling context suggests that the choral laughter does not occur randomly in the storytelling sequence but in response to the point of the storytelling, its climax, and its association with the climax is clearly evidenced by the laughter’s loudness and extended duration.

Needless to say that the XML format makes these instances of laughter, just as any other nodes in the XML network, readily available for extraction and examination. A simple XPath query to retrieve laughter occurring in overlap is this:

```
//seq[@type="ovl"]//laugh
```

This query returns all instances of simultaneous laughter with all the meta-data included in the elements. If the laughter sound, captured by the ‘desc’ attribute, is of primary interest, call the string() function:

```
//seq[@type="ovl"]//laugh[@desc]/string(@desc)
```

5. Concluding remarks

The excerpts discussed above demonstrate that corpus-linguistic transcripts and conversation-analytic transcripts can be merged and integrated into an XML architecture. Researchers will obviously need not just one such integrated XML file but large quantities. While transposing a single CA transcript into XML is already a complex task, transposing a large number of them may seem daunting. Why expend this extra effort?

The integration into XML pays dividends on three counts. Firstly, XML is by now the encoding standard for computerized text world-wide. Working on XML-formatted CA transcripts will, then, greatly enhance their exportability, storability, and sharability. Secondly, XML documents are networks in which all nodes are connected in one way or another. This omni-connectedness is key in that any node or set of nodes in the document can be addressed and extracted using appropriate XML query tools such as XPath and XQuery. For CA transcription, XML formatting is a game changer: while CA researchers have been used to searching their data manually, with all the limitations to size and extractability of data, XML transcripts allow for efficient automatic retrieval, extraction, and analysis of very complex and very large data sets. Thus, XML will also facilitate what has so far been beyond the reach of orthodox CA research: examining data with a view to large-scale quantification and eventual statistical evaluation. The third reason why the transposition into XML is worth the effort is the unique integration of two distinct approaches to conversational data — the corpus-linguistic one embodied in the exhaustive annotation for word class through POS tagging and the conversation-analytic one embodied in the careful attention to situated interactional detail. While a number of CA transcription tools such as ELAN (Wittenburg et al. 2006), EXMARaLDA (Schmidt & Wörner 2014), and FOLKER (Schmidt & Schütte 2014) do have an XML component, none of them have to date any POS tagging functionality. In other words, an integrated corpus-linguistic and conversation-analytic transcription in XML format allows high-efficiency access to conversational data both on the (lower) lexical to grammatical levels as well as the (higher) discourse and pragmatic levels of interaction — a potentially fruitful marriage in the spirit of the recent rapprochement of CL and CA witnessed in the burgeoning field of corpus pragmatics (cf., for example, Author 2015b).
### Appendix: Current tagging scheme

<table>
<thead>
<tr>
<th>Category</th>
<th>XML element</th>
<th>Sub-category</th>
<th>XML attributes &amp; attribute values</th>
<th>CA symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential aspects</td>
<td><code>&lt;seq&gt;</code></td>
<td>overlap:</td>
<td><code>&lt;seq type=&quot;ovl&quot;&gt;</code></td>
<td>[ ]</td>
<td>overlapped/overlapping speech</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;seq n=&quot;&quot;&gt;</code></td>
<td></td>
<td>number in sequence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;seq part=&quot;fst&quot; or part=&quot;scd&quot; or part=&quot;thd&quot;&gt;</code></td>
<td></td>
<td>overlapped or overlapping speech(es)</td>
</tr>
<tr>
<td></td>
<td>latching:</td>
<td></td>
<td><code>&lt;seq type=&quot;itch&quot;&gt;</code></td>
<td>=</td>
<td>one turn latched on to next turn with less-than-usual or no gap at all</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;seq part=&quot;fst&quot; or part=&quot;scd&quot; or part=&quot;thd&quot;&gt;</code></td>
<td></td>
<td>latched or latching turn(s)</td>
</tr>
<tr>
<td>Temporal aspects</td>
<td><code>&lt;tmp&gt;</code></td>
<td>pauses:</td>
<td><code>&lt;tmp type=&quot;pause&quot; dur=&quot;&quot;&gt;</code></td>
<td>(.) or (1.2)</td>
<td>short or longer pause</td>
</tr>
<tr>
<td></td>
<td></td>
<td>speed-up:</td>
<td><code>&lt;tmp spd=&quot;&lt;&quot; or spd=&quot;&lt;&lt;&quot; or spd=&quot;&lt;&lt;&lt;&quot;&gt;</code></td>
<td>&lt;&lt;</td>
<td>increase in speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>slow-down:</td>
<td><code>&lt;tmp slw=&quot;&quot;&gt; or slw=&quot;&gt;&gt;&quot; or slw=&quot;&gt;&gt;&gt;&quot;</code></td>
<td>&gt;&gt;</td>
<td>decrease in speed</td>
</tr>
<tr>
<td>Phonological</td>
<td><code>&lt;phn&gt;</code></td>
<td>intonation:</td>
<td><code>&lt;phn int=&quot;rise&quot;&gt;</code></td>
<td>?</td>
<td>question(-like) rise</td>
</tr>
<tr>
<td>aspects</td>
<td></td>
<td></td>
<td><code>&lt;phn int=&quot;halfrise&quot;&gt;</code></td>
<td>¬</td>
<td>rise stronger than a comma but weaker than a question mark</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn int=&quot;weakrise&quot;&gt;</code></td>
<td>¬,</td>
<td>weakly rising intonation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn int=&quot;fall&quot;&gt;</code></td>
<td>.</td>
<td>falling intonation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn int=&quot;ctd&quot;&gt;</code></td>
<td>,</td>
<td>continued intonation</td>
</tr>
<tr>
<td></td>
<td>pitch change:</td>
<td></td>
<td><code>&lt;phn ptch=&quot;up&quot;&gt;</code></td>
<td>↑</td>
<td>sharp rise in pitch</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn ptch=&quot;updown&quot;&gt;</code></td>
<td>↑↓</td>
<td>sharp risefall in pitch</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn ptch=&quot;down&quot;&gt;</code></td>
<td>↓</td>
<td>sharp fall in pitch</td>
</tr>
<tr>
<td></td>
<td>volume:</td>
<td></td>
<td><code>&lt;phn vol=&quot;high&quot;&gt;</code></td>
<td>A</td>
<td>loud voice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><code>&lt;phn vol=&quot;low&quot; degr=&quot;°&quot; or degr=&quot;°°&quot; or degr=&quot;°°°&quot;&gt;</code></td>
<td>°a</td>
<td>soft voice; three degrees</td>
</tr>
<tr>
<td></td>
<td>stretching:</td>
<td></td>
<td><code>&lt;phn lgth=&quot; &quot; degr=&quot;.&quot; or degr=&quot;.:&quot; or degr=&quot;::&quot;&gt;</code></td>
<td>a::</td>
<td>lengthened sound; three degrees</td>
</tr>
<tr>
<td>Category</td>
<td>XML element</td>
<td>Sub-category</td>
<td>XML attributes &amp; attribute values</td>
<td>CA symbol</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>stress:</td>
<td>&lt;phn strs=&quot;&quot; desc=&quot;much&quot; or desc=&quot;more&quot; or desc=&quot;most&quot;&quot;&gt;</td>
<td>a or a or A</td>
<td>stressed or heavily stressed or very heavily stressed sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>realization:</td>
<td>&lt;phn real=&quot;&quot;&gt;</td>
<td></td>
<td>deviant realization of (part of) word</td>
<td></td>
<td></td>
</tr>
<tr>
<td>truncation:</td>
<td>&lt;phn trnc=&quot;&quot;&gt;</td>
<td>-</td>
<td>cut-off in mid-word</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aspiration:</td>
<td>&lt;phn asp=&quot;in&quot; or asp=&quot;ex&quot;&gt;</td>
<td>.h or h.</td>
<td>inhalation or exhalation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>smile voice:</td>
<td>&lt;phn smv=&quot;£&quot;&gt;</td>
<td>£</td>
<td>talk produced while smiling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laughter</td>
<td>&lt;laugh&gt;</td>
<td>within-speech:</td>
<td>&lt;laugh type=&quot;wsl&quot; desc=&quot;&quot;&gt;</td>
<td>a(h)a</td>
<td>laughing within words; exact description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;laugh vol=&quot;high&quot; which=&quot;&quot;&gt;</td>
<td>indication of which sounds were affected by loud wsl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;laugh vol=&quot;low&quot; which=&quot;&quot;&gt;</td>
<td>indication of which sounds were affected by soft within-speech laughter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>between-speech:</td>
<td>&lt;laugh type=&quot;bsl&quot; desc=&quot;&quot; dur=&quot;&quot;&gt;</td>
<td>ha ha</td>
<td>laughing between words; exact description and duration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;laugh vol=&quot;high&quot; or vol=&quot;low&quot;&gt;</td>
<td>loud or soft between-speech laughter</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td>&lt;cmmt&gt;</td>
<td>on hearing:</td>
<td>&lt;cmmt aud=&quot;unc&quot;&gt;</td>
<td>( )</td>
<td>unclear hearing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;cmmt aud=&quot;poss&quot;&gt;</td>
<td>( a )</td>
<td>possible hearing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;cmmt aud=&quot;altn&quot; which=&quot;&quot;&gt;</td>
<td>( a / b )</td>
<td>alternative hearings; alternative is specified in ’which’ attribute</td>
</tr>
<tr>
<td></td>
<td></td>
<td>on event:</td>
<td>&lt;cmmt evnt=&quot;&quot;&gt;</td>
<td>( ( ) )</td>
<td>extra-linguistic event</td>
</tr>
<tr>
<td></td>
<td></td>
<td>on anything else:</td>
<td>&lt;cmmt oth=&quot;&quot;&gt;</td>
<td></td>
<td>other types of comment</td>
</tr>
</tbody>
</table>

References:
Author 2013
Author 2015a
Author 2015b


