

# The British Academic Spoken English Corpus Manual

## Transcription and Mark up

### 1. Transcription

Spellings have been taken from [Online Oxford Reference](#) or the [OED Online](#) or have been noted below for consistency.

#### 1.1 General:

- Oxford Online Reference was used as a first step to finding the orthographic spelling of a word.
- Where the spelling does not appear in the reference books, Google was used for suggestions and the most popular occurrence taken, adding it to the list below to maintain consistency.
- Entity references have been used for transcribing characters outside the set of characters available, using entity names rather than decimal codes i.e. &acute;, &ouml; and so on.
- The list of words that are always –ise rather than –ize are:

advertise	dis(en)franchise	merchandise
advise	disguise	prise (open)
apprise	enfranchise	revise
chastise	enterprise	supervise
circumcise	excise	surmise
comprise	exercise	surprise
demise	franchise	televisise
despise	improvise	
devise	incise	

- *Gonna, wanna, kinda* etc. are transcribed in their full forms: *going to, want to, kind of*.
- The spelling from the above references has been used even if a spelling has been used differently by a lecturer e.g. a lecturer has *Perikles* on a handout consistently, but *Pericles* was used as the spelling as it occurs in this form in the dictionary.
- Other variable spelling and words not found in the dictionary:

aagh  
adviser  
analyse  
biggie  
blebs  
bleurgh  
chock-a  
combating  
e-commerce  
egg cup  
e-mail  
encyclopedia  
et cetera  
fetus  
filestore  
filmmaker filmmaking  
flak  
focused  
frusemide  
geektastic  
grottoes  
haemorrhage  
Habsburg  
hand-wavy  
huh  
hurray - not hooray  
immunosuppression  
infrared  
in so far - not insofar  
judgement  
linchpin  
leukocyte  
mangoes  
medieval

mega market  
minijoke  
ml  
monoline  
naive  
nineteen-o-five  
no-no  
oedema  
oi  
okey-dokey  
oof  
per cent  
piccies  
scubaing  
secondborn  
shh  
so-called  
straight away  
swap  
technobabble  
thingy  
ticklist  
ton rather than tonne, if it is not clear  
trade-off  
T-test  
weblog  
web page  
web site  
whisky  
whoops and whoop  
woo-hoo  
yeah  
yep – *not yup*

## 1.2 Capitalization

- Book/film titles have been capitalised by capitalising the first word and all other content words in the title:  
The Faerie Queene  
The Adventures of Huckleberry Finn
- All units have lower case letters even if they are the names of scientists:  
newtons, daltons
- Names of departments are capitalized:  
Department of History but  
History department
- Directions are only capitalized if they form part of a proper name:  
south Cheshire  
South Yorkshire

## 1.3 Hyphenation

### 1.3.1 General

In general, hyphens are used sparingly. Conventions from Oxford Online Reference and OED Online are taken, and unless they appear explicitly, follow the conventions noted below.

- Hyphens for formulae, using the grouping for disambiguation:  
the output is  $K$  over  $\alpha$ -squared-plus- $\beta$ -squared the one-over- $S$  there it is again and two terms #  
exponential-minus- $\alpha$ - $T$  multiplied by the cosine of  $\beta$ - $T$  square-root of  $\alpha$ -squared-plus- $\beta$ -  
squared on the top divided by  $\beta$  that's that which again is just a gain component just a number and  
then we've got the two terms exponential-minus- $\alpha$ - $T$  multiplied by sine- $\beta$ - $T$
- For truncation within a formulae or other hyphenated section:  
<trunc>M-V-s</trunc> squared-over-six
- For web addresses:  
W-W-W-dot-NATO-dot-org
- For connecting non-word spoken noise components:  
dah-di-dah-di-dah  
ba-de-bum

### 1.3.2 Hyphenated

- non-something words
- pre-something is only hyphenated if the word that is pre- begins with e or i or if the formation could be ambiguous e.g. preposition/preposition
- post-something words
- something-like words
- something-related words
- something-specific words
- mid-something words
- something-shaped words
- counter-something words
- anti-something words
- quasi-something words
- socio-something words
- semi-something words
- pro-something words
- pseudo-something words

T-helper cell  
P-value  
twelve-pounds-ninety-nine  
first-hand  
un-Roman  
oft-cited  
okey-dokey  
wire-free  
arch-example  
one-legged  
touchy-feely  
two-hundred-and-something  
fifty-odd

### 1.3.3 Not hyphenated

- cosomething e.g. coworker
- resomething e.g. reread
- somethingish e.g. yellowish
- subsomething e.g. subgroup
- somethingwise e.g. personalitywise
- protosomething e.g. prototype
- missomething e.g. misemphasizing

### 1.3.4 Other:

middle-sized  
one-sided  
south-east  
T-cell

## 1.4 Anonymization

- The <gap/> element is used for anonymization.
- <gap reason="ethics" extent="N secs"/> is used for any portion of an event that is to be removed for ethical reasons or that is requested by the participants to be removed.

## 2. Mark up

Occurrences are marked up following the definitions below

### 2.1 <u>

Element	Attributes	Values	Examples
<p><b>&lt;u&gt;</b></p> <p><i>An utterance is a discrete sequence of speech produced by one participant, or group of participants, in a speech event. The tag contains transcription of lexical items. A &lt;u&gt;lasts until another speaker begins, there is a break in the recording or there is another stimulus which takes over e.g. video plays.</i></p>	<p><b>who (currently n)</b></p> <p><i>Supplies identifier for the speaker or group of speakers. Speaker ID supplied at beginning of every new turn.</i></p>	<p>6 character code unless there are multiple speakers in unison or saying the same thing not necessarily in unison if individual speakers cannot be distinguished:</p> <p><b>ss</b> audience members, not necessarily students</p> <p><b>sl</b> audience members and current speaker</p>	<p><b>&lt;u who="ss"&gt;</b> transcribed text <b>&lt;/u&gt;</b></p> <p><b>&lt;u who="om0095"&gt;</b> transcribed text <b>&lt;/u&gt;</b></p> <p><b>&lt;u who="nf1336"&gt;</b> transcribed text <b>&lt;/u&gt;</b></p> <p><b>&lt;u who="sm0574"&gt;</b> transcribed text <b>&lt;/u&gt;</b></p>
		<p><u>1<sup>st</sup></u> character:</p> <p>characterises speaker</p> <p><b>o, n</b> or <b>s</b></p> <p><b>o</b> observer i.e. cameraperson</p> <p><b>n</b> non-student</p> <p><b>s</b> student</p>	
		<p><u>2<sup>nd</sup></u> character:</p> <p>sex of speaker</p> <p><b>m, f</b> or <b>u</b></p> <p><b>m</b> male</p> <p><b>f</b> female</p> <p><b>u</b> unknown</p>	

		<u>3rd-6th character:</u> 4 digit number individual to each speaker <b>0000-9999</b> [The ID numbers (last 4 characters) consist of 0001 onwards in the lectures, and 5001 onwards in the seminars]	
	<b>trans</b> <i>Used when change of speaker occurs</i>	<b>latching</b> the utterance begins with a pause shorter than 0.2 second	<u who="nf1336" trans="latching"> transcribed text </u>
		<b>overlap</b> the utterance begins before the previous one has finished	<u who="om0095" trans="overlap"> transcribed text </u>
		<b>pause</b> the utterance begins after a pause of 0.2 second or longer	<u who="sm0574" trans="pause"> transcribed text </u>

Further examples:

- Addressing non-responsive audience

If a speaker addresses an audience with a question and there is no response, there is no break in the utterance tag, it remains the same utterance until a break in the recording, or another speaker speaks or there is a change in primary stimulus.

- Break in recording

Where there is a break in recording, the utterance ends, followed by the <gap> tag, see below. A new <u> will begin, with the trans attribute assigning a pause value.

lslct026

that's actually doubling your urine output so although per minute that doesn't seem much you take that into account over a day that's a huge volume change </u><gap reason="break in recording" extent="uncertain"/> <u n="nf0368" trans="pause"> so moving on now to renal clearance and that's simply the ability of your kidneys to clear whatever you're talking about from the blood

- Interaction

Where another person speaks, a new <u> is assigned. If the transition value between utterances is pause, where the definition of pause is the same as that for the <pause/> tag, the tag goes between </u> and <u> and the trans attribute is assigned a pause value:

pslct016

what would the demand curve have to look like <pause dur="0.7"/> for the elasticity not to be a negative number <pause dur="2.5"/><kinesic desc="indicates member of audience" iterated="n"/> point here </u><u n="sm0776" trans="latching"> when the demand is constant </u><pause dur="0.4"/> <u n="nm0775" trans="pause"> when demand is constant so let's look at that in the diagram

- Other stimulus

Where another stimulus takes over, the <u> ends. It begins again when a speaker begins.

ahlct014

okay so this is The Nation's Health it's from nineteen-eighty-two </u><pause dur="42.4"/> <event desc="starts video" iterated="n" n="nm0063"/><kinesic desc="video plays" iterated="y" dur="40"/> <u n="nm0063" trans="pause"> okay <pause dur="0.9"/><event desc="stops video" iterated="n"/> that scene actually comes after <pause dur="0.5"/> # <pause dur="1.4"/> an earlier consultation that he has with a woman

## 2.2 <pause/>

Element	Attributes	Values	Examples
<b>&lt;pause/&gt;</b> <i>Indicates a pause, either between or within utterances. A pause is a silence in speech and vocality of the speaker for 0.2 second or above.</i>	<b>dur</b> <i>duration of pause</i>	measurement of 0.2 second or above, correct to 1 decimal place	the whole environment around <b>you</b> <b>&lt;pause dur="0.3"/&gt;</b> matches <b>&lt;pause dur="0.2"/&gt;</b> your mood <b>&lt;pause dur="2.0"/&gt;</b> now

- Pauses are breaks in speech or any vocal noise made by the speaker. Laughter, coughs, clearings of throat etc. are not contained within the pause measurement if they are made by the speaker or the speaker with other participants where the who attribute contains “I”. Where there are vocals from other participants in the event, these are included in the pause measurement while it occurs within the <u> of the current speaker.

sslct009

i'll be in my room during <pause dur="0.2"/> <trunc>s</trunc> so-called <pause dur="0.4"/> surgery hours <pause dur="0.4"/><vocal desc="laugh" iterated="n"/><pause dur="1.4"/> right let's get this thing moving

The <vocal/> here is not contained within the pause measurement, as it is made by the current speaker. The measurements would be the same if the who/n attribute contained “I”, e.g. “sI”.

ahlct016

the little bit from Grisbi which you did not see this morning thanks to the British censor of the fifties <pause dur="0.5"/> <vocal desc="laughter" iterated="y" n="ss" dur="1"/> and <pause dur="0.3"/> that is when # <pause dur="0.2"/> Max takes his friend Riton to # <pause dur="0.6"/> his new flat

The <vocal/> in this instance is contained within the pause measurement as it is made by “ss”, other participants in the lecture.

- Wavesurfer was used to measure pauses from the waveform of the audio.
- Pause measurements do not contain filled pauses which are transcribed as #.

lslct024

it'll fill a little gap in the agenda <pause dur="0.6"/> # <pause dur="0.2"/> just to give you some history

- <pause/> tags can occur within a word.
- Where there is overlapping speech, pauses are measured within the <u> of one speaker to their own next utterance.

## 2.3 <vocal/>

Element	Attributes	Values	Examples
<b>&lt;vocal/&gt;</b> <i>A non-lexical vocal event such as laughter, coughing. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	<b>desc</b> <i>Description of vocal event.</i>	<b>audible inbreath</b> <b>belch</b> <b>blows nose</b> <b>buzzing noise</b> <b>car stalling noise</b> <b>cough</b> <b>clears throat</b> <b>exclamation</b> <b>exploding noise</b> <b>gasp</b> <b>hum</b> <b>kiss noise</b> <b>laugh</b> <b>laughter</b> <b>shudder noise</b> <b>sigh</b>	<b>&lt;vocal desc="clears throat" iterated="n"/&gt;</b>  <b>&lt;vocal desc="laughter" iterated="y" n="ss" dur="1"/&gt;</b>



		<b>sneeze</b> <b>sniff</b> <b>stretch noise</b> <b>throat</b> - creakiness in throat <b>whistle</b> <b>whooshing noise</b> <b>wrong answer noise</b> ...	
	<b>iterated</b>	<b>n</b> if not continuous/iterated <b>y</b> if continuous/iterated	
	<b>who</b>	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	<b>dur</b> <i>duration of &lt;vocal/&gt; to nearest second. Only measure if iterated = "y"</i>	measurement of <vocal/> to nearest second.	

- All vocals are shown to the nearest word. Words are not split by <vocal/> tags.

## 2.4 <kinesic/>

Element	Attributes	Values	Examples
<p><b>&lt;kinesic/&gt;</b>  <i>A non-vocal communicative event such as put hand up, frown, act of writing on board etc. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i></p>	<p><b>desc</b>  <i>Description of kinesic.</i></p>	<p><b>changes</b> (transparency   slide)  <b>indicates</b> (board on wall   discs   line on transparency   member of audience   point on screen   point on flip chart   point on slide   point on transparency   pole   screen   transparency   section of board   upwards   point on stomach   camera   centre of book   indicator   basket   fire exit   kidneys   and counts students   size with hands   point on arm)  <b>puts</b> (petri dish on overhead projector   on transparency   hand over mouth   hand to ear   hands on head   hand round throat   transparency on top of current transparency   finger up)  <b>reveals covered</b> (part of transparency   phrase on board)  <b>moves</b> (chess pieces to memorized positions   discs   disc   hands to side of face demonstrating blinkers   hands round head demonstrating mortar board)  <b>overhead projector is on showing</b> ( transparency   covered transparency)  <b>projector is on showing slide</b>  <b>turns on overhead projector showing transparency</b>  <b>turns on projector showing slide</b>  <b>starts stopwatch</b>  <b>writes on</b> (board   transparency   flip chart)</p>	<p><b>&lt;kinesic desc="changes slide" iterated="n"/&gt;</b></p> <p><b>&lt;kinesic desc="indicates point on slide" iterated="n"/&gt;</b></p> <p><b>&lt;kinesic desc="changes transparency" iterated="y" dur="10"/&gt;</b></p> <p><b>&lt;kinesic desc="reveals covered part of transparency" iterated="n"/&gt;</b></p> <p><b>&lt;kinesic desc="writes on board" iterated="y" dur="14"/&gt;</b></p> <p><b>&lt;kinesic desc="turns on overhead projector showing transparency" iterated="n"/&gt;</b></p> <p><b>&lt;kinesic desc="demonstrates giraffe bending down" iterated="n"/&gt;</b></p> <p><b>&lt;kinesic desc="applause" iterated="y" n="ss" dur="6"/&gt;</b></p>

		<p><b>holds up</b> (reel of cotton thread   report   sheet   sheet music   bar code   prop   fist   biscuits   book   magazine   poster   ten pound note   strip of paper with weight attached   metal chain   metal chain with weight attached)</p> <p><b>demonstrates</b> (motion on transparency   motion with arms   beam with pen   giraffe bending down   big eyes with hands   head length   tube lengthening   carrying heavy object   Belousov-Zhabotinsky reaction   rejection of face   pretend breathalyser)</p> <p><b>adjusts</b> (transparency   projector)</p> <p><b>makes</b> (questioning gesture   quotation mark gesture)</p> <p><b>put hands up</b></p> <p><b>nod heads</b></p> <p><b>raises eyebrows</b></p> <p><b>shrugs shoulders</b></p> <p><b>waves to member of audience</b></p> <p><b>crosses arms across body</b></p> <p><b>applause</b></p> <p><b>clicks fingers</b></p> <p><b>gestures stop with hands</b></p> <p><b>memorizes board position</b></p> <p><b>mixes up pieces on transparency</b></p> <p><b>presses finger into arm</b></p> <p><b>sits back in chair</b></p> <p><b>brings out prop</b></p> <p><b>removes discs from poles</b></p> <p><b>adds disc to pole</b></p> <p><b>sticks bar code on sheet</b></p> <p><b>turns strip of paper with weight attached the other way up</b></p> <p><b>throws away ten pound note</b></p> <p><b>waves arms and shakes body</b></p> <p><b>starts video</b></p>	
--	--	--	--

		<b>video plays</b> <b>pulls apart piece of Velcro</b> <b>hold both hands and lean back</b> <b>displays models</b> <b>mimics falling asleep</b> <b>tilts book</b> <b>hits book</b> <b>drops book</b> <b>knocks on table</b> ...	
	<b>iterated</b>	<b>n</b> if not continuous/iterated <b>y</b> if continuous/iterated	
	<b>who</b>	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	<b>dur</b> <i>duration of &lt;kinesic/&gt; to nearest second. Only measure if iterated = "y"</i>	measurement of <kinesic/> to nearest second. If it is over a minute, the notation minute:seconds e.g. dur="3:20" <b>unknown</b> is used if duration is not known.	

## 2.5 <event/>

Element	Attributes	Values	Examples
<p><b>&lt;event/&gt;</b>  <i>An occurrence, not necessarily communicative, usually non-verbal, noted because it affects comprehension of the surrounding discourse.</i></p>	<p><b>desc</b>  <i>Description of event</i></p>	<p><b>walks</b> (to front of class   to position   to seat)  <b>takes</b> (off transparency   off jumper   off glasses   handout from member of audience   off transparency backing paper)  <b>looks through</b> (transparencies   notes)  <b>puts</b> (on blank transparency   on transparency   video into video player   pen in bin   sheets on board   away screen   away projector   on glasses   on transparency backing paper)  <b>covers part of transparency</b>  <b>knock on door, student enters</b>  <b>students enter room</b>  <b>moves</b> (board   television   blinds)  <b>wipes board</b>  <b>phone</b> (beeps   rings)  <b>noise from</b> (computer   audience)  <b>prepares video</b>  <b>rewinds video</b>  <b>fast-forwards video</b>  <b>stops video</b>  <b>turns</b> (on/off overhead projector   on/off television   turns on overhead projector, bulb blows   on/off lights)  <b>adjusts overhead projector</b>  <b>claps hands</b>  <b>passes out</b> (handouts   attendance sheet   handout to observer)  <b>pulls down screen</b></p>	<p>is there # something  i'm not seeing &lt;pause dur="0.4"/&gt;  <b>&lt;event desc="opens drawer" iterated="n"/&gt;</b> aha</p> <p>right i should get these lights out  &lt;pause dur="3.6"/&gt;<b>&lt;event desc="turns off lights" iterated="n"/&gt;</b> okay</p> <p>the focus isn't right is it &lt;pause dur="2.2"/&gt;<b>&lt;event desc="adjusts overhead projector" iterated="n"/&gt;</b> that's better</p> <p><b>&lt;event desc="looks through transparencies" iterated="y" dur="11"/&gt;</b> right when i &lt;pause dur="1.0"/&gt; finally find &lt;pause dur="1.8"/&gt; the relevant slides</p>

		<p><b>presses light switches</b>  <b>changes</b> (lights   pen   tape   pointer)  <b>drops</b> (pen   pencil case   radio microphone transmitter   board rubber)  <b>shuts curtains</b>  <b>removes jacket</b>  <b>woman enters room, gives the lecturer an attendance sheet</b>  <b>closes computer program</b>  <b>replaces transmitter on belt</b>  <b>picks up piece of radio microphone from the floor</b>  <b>knocks microphone off</b>  <b>tries to</b> (reattach microphone   show slides)  <b>moves</b> (microphone closer to mouth   to slide room   away from the overhead projector light)  <b>returns to position</b>  <b>discussing questions</b>  <b>hits surface</b>  <b>hands prop to student</b>  <b>looks for board rubber</b>  <b>stands on equipment</b>  <b>accelerating beeps played</b>  <b>tries to change slide</b>  <b>holds up hands demonstrating tentacles in front of face</b>  <b>gets equipment</b>  <b>attempts task set</b>  <b>doing task set</b>  <b>drinks</b>  <b>stops notes from falling off table</b>  <b>finds pointer</b>  <b>takes</b> (handout from member of audience   off transparency backing paper)  <b>attaches weight to metal chain</b></p>	
--	--	---	--

		<b>music plays</b> <b>shakes liquid</b> <b>taps microphone</b> <b>opens drawer</b> <b>looks at timetable</b> <b>overhead projector is on showing</b> <b>blank transparency</b> <b>beep from stopwatch</b> ...	
	<b>iterated</b>	<b>n</b> if not continuous/iterated <b>y</b> if continuous/iterated	
	<b>who</b>	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	<b>dur</b> <i>duration of &lt;event/&gt; to nearest second. Only measure if iterated = "y"</i>	measurement of <event/> to nearest second. If it is over a minute, use the notation minute:seconds e.g. dur="3:20" <b>unknown</b> if duration is not known.	

## 2.6 <shift/>

Element	Attributes	Values	Examples
<b>&lt;shift/&gt;</b> <i>A marked change in voice quality for any one speaker</i>	<b>feature</b> <i>the paralinguistic feature</i>	<b>tempo</b> speed of utterance <b>loud</b> loudness <b>pitch</b> pitch range <b>tension</b> tension or stress pattern <b>rhythm</b> rhythmic qualities <b>voice</b> voice quality	many of the ancient myths that you should be coming across in your classical reading <pause dur="0.3"/> <b>&lt;shift feature="voice" new="laugh"/&gt;</b> show that no such thing <b>&lt;shift feature="voice" new="normal"/&gt;</b> <pause dur="0.3"/> is the case  Sainsbury's goes down threepence <pause dur="0.4"/> on # baked beans <b>&lt;shift feature="voice" new="mimicking an angry voice"/&gt;</b> you bastard <pause dur="0.2"/>
	<b>new</b> <i>new state of paralinguistic feature</i>	for "voice" <b>laugh</b> laughing <b>whisp</b> whispering <b>husk</b> husky <b>growl</b> growling voice <b>mimicking French accent</b> <b>mimicking Jamaican accent</b> <b>mimicking northern English</b>	

		<p> <b>accent</b>  <b>mimicking American accent</b>  <b>mimicking Essex accent</b>  <b>mimicking Australian accent</b>  <b>mimicking an other's voice</b>  <b>mimicking an angry voice</b>  <b>mimicking an orgasmic voice</b>  <b>normal</b>  ...  for "pitch"  <b>high</b>  <b>normal</b>  ...  see  <a href="http://www.tei-c.org/P4X/TS.html#TSSASH">http://www.tei-c.org/P4X/TS.html#TSSASH</a> </p>	<pre> fourpence on ravioli you swine &lt;shift feature="voice" new="normal"/&gt;frozen peas down &lt;pause dur="0.4"/&gt;&lt;vocal desc="laughter" iterated="y" n="ss" dur="1"/&gt; and who the hell wins  swing it around the crowd and &lt;shift feature="pitch" new="high"/&gt;ding &lt;shift feature="pitch" new="normal"/&gt;&lt;pause dur="0.4"/&gt;&lt;kinesic desc="indicates member of audience" iterated="n"/&gt; it picks you out </pre>
--	--	---	---

- The shift tag is used sparingly as it is imprecise and difficult to maintain consistency. It is not used for marking prosody generally, only when it is a section of speech that can be classified easily as it is produced differently by speaker.
- See <http://www.tei-c.org/P4X/TS.html#TSSASH> for more coding for <shift/> tag



## 2.7 <distinct>

Element	Attributes	Values	Examples
<b>&lt;distinct/&gt;</b> <i>Used for words or phrases in languages other than present-day British English. This includes earlier forms of English but does not include proper names. Foreign words are spelled as in the original language when it uses a roman alphabet; otherwise, a standard form of romanization is used</i>	<b>lang (currently n)</b> <i>duration of pause</i>	2/3 letter code as defined in Library of Congress <a href="http://www.loc.gov/standards/iso639-2/langcodes.html">http://www.loc.gov/standards/iso639-2/langcodes.html</a>  <b>cel</b> Celtic <b>fr</b> French <b>de</b> German <b>grc</b> Ancient Greek <b>el</b> Modern Greek <b>it</b> Italian <b>is</b> Icelandic <b>ja</b> Japanese <b>la</b> Latin <b>no</b> Norwegian <b>oe</b> Old English <b>po</b> Polish <b>ru</b> Russian <b>sa</b> Sanskrit <b>sco</b> Scots <b>es</b> Spanish <b>sw</b> Swahili/Kiswahili ...	a kind of <pause dur="0.4"/> civil service <trunc>fonc</trunc> <b>the</b> <distinct n="fr">fonction publique</distinct>  translation of true and fair view <pause dur="0.5"/> <distinct n="el">di pragmatiki eikona</distinct> <pause dur="0.4"/> <distinct n="fr">une image fid&egrave;le</distinct> <pause dur="0.4"/> <distinct n="es">la</distinct> <pause dur="0.2"/> do you remember the Spanish case <pause dur="0.5"/> <distinct n="es">la imagen fiel</distinct> <pause dur="0.8"/> the you don't remember this i talked about it last time  is like this <pause dur="3.9"/> # <pause dur="0.2"/> <distinct n="sa">janasas</distinct> i don't know how you pronounce it <distinct n="sa" type="sampa">[dZ{n{s{s}</distinct> or something like that  there's a <distinct type="sampa">[r]</distinct> and there's a <distinct type="sampa">[s]</distinct> <pause dur="0.6"/> but in Greek there is nothing
	<b>type</b> <i>type is used when the transcription changes, for example, when it is the pronunciation of the word that affects the comprehension of the discourse.</i>	<b>sampa</b> ...	

## 2.8 <sic>

Element	Attributes	Values	Examples
<b>&lt;sic&gt;</b> <i>Used when a speaker makes a mistake without self-correcting, and the error might otherwise appear to be a transcribing error.</i>	<b>corr</b> <i>correction</i>	...	there are <trunc>th</trunc> <b>&lt;sic corr="really"&gt;</b> beally <b>&lt;/sic&gt;</b> three <pause dur="0.2"/> parts to <pause dur="0.4"/> the poster

- Used sparingly for uncorrected slips of the tongue and only if it could be seen as a transcribing error.
- Not used to correct grammar as with written text.
- Used for a truncated word followed by correction if not all of the word is said:  
 the environment opposed **<sic corr="imposed">**impo-, posed-**</sic>** upon them

## 2.9 <trunc>

Element	Attributes	Values	Examples
<b>&lt;trunc&gt;</b> <i>Used when a word is truncated</i>			is that i think <b>&lt;trunc&gt;</b> w <b>&lt;/trunc&gt;</b> <pause dur="0.5"/> today's session is going to be <pause dur="0.3"/> the last  with other <pause dur="0.2"/> <b>&lt;trunc&gt;</b> poin <b>&lt;/trunc&gt;</b> other <pause dur="0.2"/> bodies

- Truncation tags are used when there is a truncation at the beginning of the word as well as if there is a truncation at the end of the word.

## 2.10 <gap/>

Element	Attributes	Values	Examples
<b>&lt;gap&gt;</b> <i>Used to indicate omissions in the text and recording. Also used when names referred to in the recording are withheld.</i>	<b>reason</b>	<b>inaudible</b> <b>inaudible due to noise from audience</b> - includes coughs, etc <b>inaudible due to equipment difficulties</b> <b>inaudible due to hand over mouth</b> <b>inaudible, multiple speakers</b> <b>inaudible due to overlap ethics</b> e.g. "i hope this will be removed" <b>name</b> ...	C-D-four cells undergo a <pause dur="0.6"/> diurnal rhythm <pause dur="0.2"/> that is to say they're there's a </u><gap reason="break in recording" extent="uncertain"/> <u n="nm0239" trans="pause"> if you measure them <pause dur="0.4"/> at different times of the day you get different numbers  i think we'd better get started <pause dur="1.0"/> i'm <gap reason="name" extent="2 words"/> and i'm just giving a lecture # <pause dur="0.4"/> today on <pause dur="0.2"/> Max Weber  it can pick up electrons <pause dur="0.2"/> to give this <pause dur="0.3"/><kinesic desc="indicates point on transparency" iterated="n"/> <gap reason="inaudible" extent="1 sec"/> that's E-minus of the proton is equivalent of H atom
	<b>extent</b>	<b>N words</b> <b>N secs</b> <b>uncertain</b>	

## 2.11 <unclear>

Element	Attributes	Values	Examples
<b>&lt;unclear&gt;</b> <i>Used when transcriber is uncertain of exact word(s)</i>			and you've also got the stirrer <b>&lt;unclear&gt;</b> wheel <b>&lt;/unclear&gt;</b> <pause dur="0.5"/> okay so that that would be a typical batch thing

## 2.12 <reading>

Element	Attributes	Values	Examples
<b>&lt;reading&gt;</b> <i>text which can be attributed to an identifiable source when it is being read and not referenced where the whole text being read is at the non-finite clause and above level.</i>			though how far it was the norm <pause dur="1.2"/> # <pause dur="1.6"/> leaves certain questions in our minds <pause dur="0.3"/> he says that # <pause dur="1.1"/> <b>&lt;reading&gt;</b> Agricola was no loose young<unclear>subbleton</unclear> <pause dur="0.4"/> to turn his military career into a life of gaiety he wouldn't make his self-captaincy <pause dur="0.3"/> <trunc>as</trunc> and his inexperience an excuse for idly enjoying himself <pause dur="0.3"/> and continually going on leave <b>&lt;/reading&gt;</b> <pause dur="0.3"/> giving us <pause dur="0.3"/> the impression that the norm <pause dur="0.4"/> for <pause dur="0.2"/> senior tribunes <pause dur="0.6"/> was basically to use <pause dur="0.4"/> their period <pause dur="0.3"/> of supposed military service <pause dur="0.2"/> as a long <pause dur="0.2"/> and extended holiday

## 2.13 Further examples

- Using audio

Marking audio uses three tags: `<event desc="starts audio" n="nm0858" iterated="n"/>`  
`<kinesic desc="audio plays" iterated="y" dur="13"/>`  
`<event desc="stops audio" n="nm0858" iterated="n"/>`

which can be placed either outside `<u>` tags or within them. The `<u>` tags are used for coding the speakers and the placement of the tags marking the events also mark the temporal placement of the actions. The `n/who` attribute marks who is doing the action if the tag is outside a `<u>`.

pslct024

you can recognize the word every single time `</u>` `<pause dur="19.6"/>``<event desc="starts audio"`  
`n="nm0858" iterated="n"/>``<kinesic desc="audio plays" iterated="y" dur="13"/>` `<event desc="stops audio"`  
`n="nm0858" iterated="n"/>` `<u n="nm0858" trans="pause">` it goes on for hours

- Using video

Marking video uses three tags: `<event desc="starts video" n="nf0058" iterated="n"/>`  
`<kinesic desc="video plays" iterated="y" dur="3:01"/>`  
`<event desc="stops video" iterated="n"/>`

which can be placed either outside `<u>` tags or within them. The `<u>` tags are used for coding the speakers and the placement of the tags marking the events also mark the temporal placement of the actions. The `n/who` attribute marks who is doing the action if the tag is outside a `<u>`.

- Demonstrating Towers of Hanoi puzzle

pslct010

this is my `<pause dur="1.3"/>``<kinesic desc="brings out prop" iterated="n"/>` i i `<pause dur="0.7"/>`  
cheated earlier today `<pause dur="0.5"/>` because i # `<pause dur="0.8"/>` i pretended to be a student  
`<pause dur="0.6"/>` in the # `<pause dur="0.6"/>` writing course that that `<pause dur="0.6"/>` C-S take  
`<pause dur="1.0"/>` # and there were lots of exciting things happening in `<pause dur="0.6"/>` in that  
there was music being played and `<pause dur="1.0"/>` tenners being thrown up scrunched into balls and  
thrown `<pause dur="0.6"/>` into the corner `<pause dur="1.1"/>` and it continues `<pause dur="0.2"/>` the  
excitement of the day `<pause dur="0.8"/>` 'cause we have a `<pause dur="2.6"/>` a prop `<pause dur="1.8"/>`  
which is my `<pause dur="2.0"/>` Towers of Hanoi `<pause dur="1.4"/>` okay `<pause dur="0.4"/>` Towers of  
Hanoi `<pause dur="0.5"/>` three poles `<pause dur="0.6"/>` with `<pause dur="0.4"/>` a number of discs `<pause`  
`dur="0.9"/>` on one end `<pause dur="0.3"/>` and the idea is to take `<kinesic desc="indicates discs"`  
`iterated="n"/>` these three discs `<pause dur="0.2"/>` and to move them all `<pause dur="0.5"/>` to the other  
end `<pause dur="1.7"/>``<kinesic desc="indicates pole" iterated="n"/>` okay `<pause dur="0.6"/>` third pole  
`<pause dur="2.6"/>` by placing taking one at a time `<pause dur="0.5"/>``<kinesic desc="moves disc"`  
`iterated="n"/>` moving them across `<pause dur="0.3"/>` but not putting a smaller disc `<pause dur="2.4"/>` a

larger disc onto a smaller disc <pause dur="0.4"/> okay so <pause dur="0.4"/><kinesic desc="moves disc" iterated="n"/> that would be an illegal move in the Towers of Hanoi problem <pause dur="0.2"/> okay we can't do that <pause dur="0.2"/> we can only move <kinesic desc="moves discs" iterated="y" dur="5"/> smaller discs onto <pause dur="0.5"/> larger discs <pause dur="1.7"/> okay so we could solve it in various ways it could be <pause dur="0.5"/> quite tricky could be difficult

...

how many people can do it with three discs <pause dur="2.2"/><kinesic desc="put hands up" iterated="n" n="ss"/> so <pause dur="0.7"/><kinesic desc="adds disc to pole" iterated="n"/><event desc="hands prop to student" iterated="n"/> this is when i get someone to do it <pause dur="0.2"/> who did you did you put your hand up </u><pause dur="0.4"/> <u n="sm0719" trans="pause"> yeah </u><pause dur="0.3"/> <u n="nm0718" trans="pause"> go for it <pause dur="1.0"/> they're they're not in the right order by the way </u><u n="sm0719" trans="overlap"> yeah </u><u n="nm0718" trans="overlap"> so you can sort them out first <pause dur="4.4"/><kinesic desc="moves discs" iterated="y" n="sm0719" dur="30"/> okay so </u><pause dur="5.2"/> <u n="ss" trans="pause"> <gap reason="inaudible, multiple speakers" extent="15 sec"/></u><u n="nm0718" trans="overlap"> <trunc>st</trunc> <pause dur="2.6"/> you need to <trunc>st</trunc> <pause dur="0.6"/> right <vocal desc="laughter" iterated="y" n="sm0719" dur="1"/> <pause dur="0.3"/> you need to start with <trunc>st</trunc> <trunc>st</trunc> <pause dur="0.3"/><kinesic desc="moves discs" iterated="y" dur="5"/> shall we start again </u><u n="sm0719" trans="latching"> yeah <pause dur="0.2"/> go on </u><u n="nm0718" trans="overlap"> i dare you </u><u n="sm0719" trans="latching"> let's go try again </u><pause dur="1.1"/> <u n="nm0718" trans="pause"> start with them at one end <pause dur="0.8"/><kinesic desc="moves discs" iterated="y" n="sm0719" dur="29"/> that might be easier </u><pause dur="1.2"/> <u n="ss" trans="pause"> <gap reason="inaudible, multiple speakers" extent="23 sec"/><kinesic desc="applause" iterated="y" n="ss" dur="4"/></u><u n="nm0718" trans="overlap"> well done <pause dur="2.5"/> thank you <pause dur="0.8"/> # <pause dur="1.3"/> it it's <pause dur="0.6"/> that that was i'm i'm <pause dur="0.2"/> dead impressed with that actually <pause dur="0.2"/> because it's very difficult doing it when there are lots of people watching

- The first extract shows a general way to tackle this particular problem. The difference between iterated="y" and iterated="n" is that where the action is not iterated, one disc is being moved to another pole and where the action is iterated, more than one disc is being moved and the duration of seconds is noted.
- The second extract is further on in the lecture where the lecturer hands the prop to a student to attempt the puzzle.