

The British Academic Spoken English Corpus Manual

The British Academic Spoken English (BASE) corpus was developed at the Universities of Warwick and Reading, under the directorship of Hilary Nesi (Centre for English Language Teacher Education, Warwick), with Paul Thompson (Department of Applied Linguistics, Reading). Natalie Snodgrass (Warwick) and Sarah Creer (Reading) were employed as Research Assistants and Tim Kelly was Video Director for the project. Lou Burnard (Oxford University) and Adam Kilgarriff (Lexicography MasterClass Ltd) acted as consultants.

The BASE corpus is a collection of transcripts of lectures and seminars recorded at two universities in the UK during the period 1998-2005. It consists of 160 lectures and 39 seminars recorded in a variety of university departments. Holdings are distributed across four broad disciplinary groups, each represented by 40 lectures and 10 seminars.

These groups are:

- Arts and Humanities
- Life and Medical Sciences
- Physical Sciences
- Social Studies and Sciences.

The lectures and seminars have been transcribed and annotated using a system devised in accordance with the TEI Guidelines. The corpus files are in XML, and there is a DTD file which must be kept in the same folder as the corpus files, named 'base.dtd'. File names are made up of five letters and three digits, in which the first two letters indicate the disciplinary group, the next three indicate whether the file is a transcript of a lecture or a seminar, and the digits are unique identifiers:

ah [Arts and Humanities]	lct	0nn
ls [Life and Medical Sciences]	[lecture]	
ps [Physical Sciences]	sem	
ss [Social Studies and Sciences]	[seminar]	

Each file consists of a header and a body. The header contains metatextual data, and the information in the header is also contained in an Excel spreadsheet which is provided with the corpus: **BASE corpus holdings.xls**. The file is provided in Read-only format, and the information in it should not be altered in any way, although users may choose to alter the layout and the sort options.

The level of mark up in the lecture files is much higher than it is in the seminar files. The reason for this is that the seminar data is difficult to mark up consistently as the seminars contain more interactivity. The use of body language, of intonation, etc, to regulate interaction in the seminars is potentially so complex that it is difficult to assess the level of mark up that is required, and thus it was decided to keep the amount of detail at a low level and to leave the close coding of the data to the interested researcher. For all of the seminar recordings and many of the lectures, video recordings can be accessed by application to the BASE team through the BASE website (<http://www2.warwick.ac.uk/fac/soc/celte/research/base/>). The audio recordings (in MP3 format) can also be accessed.

The British Academic Written English (BASE) corpus is freely available to researchers who agree to the following conditions:

1. Corpus holdings should not be reproduced in full for a wider audience/readership (ie for publication or for teaching purposes), although researchers are free to quote short passages of text up to 100 running words, with a total of 200 running words from any given assignment
2. No part of the corpus holdings should be reproduced in teaching materials intended for publication (in print or via the internet)
3. The corpus developers should be informed of all presentations and publications arising from analysis of the corpus

Researchers must acknowledge their use of the BASE corpus using the following form of words:

The recordings and transcriptions used in this study come from the British Academic Spoken English (BASE) corpus (<http://www2.warwick.ac.uk/fac/soc/celte/research/base/>). The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi (Warwick) and Paul Thompson (Reading). Corpus development was assisted by funding from the Universities of Warwick and Reading, BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

The lecture portion of the BASE corpus can be accessed through the corpus analysis interface, Sketch Engine. All 160 lectures are included, with 40 for each general disciplinary domain. This interface allows the user to view concordance lines, form complex queries, collect word frequency data (including word lists) and more. The service requires a subscription - for details, visit the Sketch Engine website at <http://corpora.sketchengine.co.uk/auth/> . The service can be obtained initially on a 30 day trial subscription with full access to all resources.

The following sections explain the transcription and mark-up conventions used in the corpus. For further information or guidance, contact the BASE team through Paul Thompson (p.a.thompson@reading.ac.uk) or Hilary Nesi (H.J.Nesi@warwick.ac.uk).

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. ´, ö 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 $K \over \alpha\text{-squared-plus-beta-squared} \frac{\text{the one-over-S}}{\text{there it is again and two terms \#}}$
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

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
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

1.4 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><u></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 <u>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>who (currently n)</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>ss audience members, not necessarily students</p> <p>sl audience members and current speaker</p>	<p><u who="ss"> transcribed text </u></p> <p><u who="om0095"> transcribed text </u></p> <p><u who="nf1336"> transcribed text </u></p> <p><u who="sm0574"> transcribed text </u></p>
		<p><u>1st</u> character:</p> <p>characterises speaker</p> <p>o, n or s</p> <p>o observer i.e. cameraperson</p> <p>n non-student</p> <p>s student</p>	
		<p><u>2nd</u> character:</p> <p>sex of speaker</p> <p>m, f or u</p> <p>m male</p> <p>f female</p> <p>u unknown</p>	

		<u>3rd-6th character:</u> 4 digit number individual to each speaker 0000-9999 [The ID numbers (last 4 characters) consist of 0001 onwards in the lectures, and 5001 onwards in the seminars]	
	trans <i>Used when change of speaker occurs</i>	latching the utterance begins with a pause shorter than 0.2 second	<u who="nf1336" trans="latching"> transcribed text </u>
		overlap the utterance begins before the previous one has finished	<u who="om0095" trans="overlap"> transcribed text </u>
		pause 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
<pause/> <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>	dur <i>duration of pause</i>	measurement of 0.2 second or above, correct to 1 decimal place	the whole environment around you <pause dur="0.3"/> matches <pause dur="0.2"/> your mood <pause dur="2.0"/> 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 “1”, e.g. “s1”.

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
<code><vocal/></code> <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>	desc <i>Description of vocal event.</i>	audible inbreath belch blows nose buzzing noise car stalling noise cough clears throat exclamation exploding noise gasp hum kiss noise laugh	<code><vocal desc="clears throat" iterated="n"/></code> <code><vocal desc="laughter" iterated="y" n="ss" dur="1"/></code>

		laughter shudder noise sigh sneeze sniff stretch noise throat - creakiness in throat whistle whooshing noise wrong answer noise ...	
	iterated	n if not continuous/iterated y if continuous/iterated	
	who	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	dur <i>duration of <vocal/> 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><kinesic/> <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>desc <i>Description of kinesic.</i></p>	<p>changes (transparency slide) indicates (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) puts (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) reveals covered (part of transparency phrase on board) moves (chess pieces to memorized positions discs disc hands to side of face demonstrating blinkers hands round head demonstrating mortar board) overhead projector is on showing (transparency covered transparency) projector is on showing slide turns on overhead projector showing transparency turns on projector showing slide starts stopwatch writes on (board transparency flip chart)</p>	<p><kinesic desc="changes slide" iterated="n"/></p> <p><kinesic desc="indicates point on slide" iterated="n"/></p> <p><kinesic desc="changes transparency" iterated="y" dur="10"/></p> <p><kinesic desc="reveals covered part of transparency" iterated="n"/></p> <p><kinesic desc="writes on board" iterated="y" dur="14"/></p> <p><kinesic desc="turns on overhead projector showing transparency" iterated="n"/></p> <p><kinesic desc="demonstrates giraffe bending down" iterated="n"/></p> <p><kinesic desc="applause" iterated="y" n="ss" dur="6"/></p>

		<p>holds up (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>demonstrates (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>adjusts (transparency projector)</p> <p>makes (questioning gesture quotation mark gesture)</p> <p>put hands up</p> <p>nod heads</p> <p>raises eyebrows</p> <p>shrugs shoulders</p> <p>waves to member of audience</p> <p>crosses arms across body</p> <p>applause</p> <p>clicks fingers</p> <p>gestures stop with hands</p> <p>memorizes board position</p> <p>mixes up pieces on transparency</p> <p>presses finger into arm</p> <p>sits back in chair</p> <p>brings out prop</p> <p>removes discs from poles</p> <p>adds disc to pole</p> <p>sticks bar code on sheet</p> <p>turns strip of paper with weight attached the other way up</p> <p>throws away ten pound note</p> <p>waves arms and shakes body</p> <p>starts video</p>	
--	--	--	--

		video plays pulls apart piece of Velcro hold both hands and lean back displays models mimics falling asleep tilts book hits book drops book knocks on table ...	
	iterated	n if not continuous/iterated y if continuous/iterated	
	who	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	dur <i>duration of <kinesic/> 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" unknown is used if duration is not known.	

2.5 <event/>

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

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

		music plays shakes liquid taps microphone opens drawer looks at timetable overhead projector is on showing blank transparency beep from stopwatch ...	
	iterated	n if not continuous/iterated y if continuous/iterated	
	who	<i>as defined above. Speaker ID not marked if produced by current speaker, marked if not current speaker.</i>	
	dur <i>duration of <event/> 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" unknown if duration is not known.	

2.6 <shift/>

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

		<p> accent mimicking American accent mimicking Essex accent mimicking Australian accent mimicking an other's voice mimicking an angry voice mimicking an orgasmic voice normal ... for "pitch" high normal ... see http://www.tei-c.org/P4X/TS.html#TSSASH </p>	<pre> fourpence on ravioli you swine <shift feature="voice" new="normal"/>frozen peas down <pause dur="0.4"/><vocal desc="laughter" iterated="y" n="ss" dur="1"/> and who the hell wins swing it around the crowd and <shift feature="pitch" new="high"/>ding <shift feature="pitch" new="normal"/><pause dur="0.4"/><kinesic desc="indicates member of audience" iterated="n"/> 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
<p><distinct/> <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></p>	<p>lang (currently n) <i>duration of pause</i></p>	<p>2/3 letter code as defined in Library of Congress http://www.loc.gov/standards/iso639-2/langcodes.html</p> <p>cel Celtic fr French de German grc Ancient Greek el Modern Greek it Italian is Icelandic ja Japanese la Latin no Norwegian oe Old English po Polish ru Russian sa Sanskrit sco Scots es Spanish sw Swahili/Kiswahili</p>	<p>a kind of <pause dur="0.4"/> civil service <trunc>fonc</trunc> the <distinct n="fr">fonction publique</distinct></p> <p>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</p> <p>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</p> <p>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</p>
	<p>type <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></p>	<p>sampa ...</p>	

2.8 <sic>

Element	Attributes	Values	Examples
<sic> <i>Used when a speaker makes a mistake without self-correcting, and the error might otherwise appear to be a transcribing error.</i>	corr <i>correction</i>	...	there are <trunc>th</trunc> <sic corr="really"> beally </sic> 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
<trunc> <i>Used when a word is truncated</i>			is that i think <trunc> w </trunc> <pause dur="0.5"/> today's session is going to be <pause dur="0.3"/> the last with other <pause dur="0.2"/> <trunc> poin </trunc> 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
<gap> <i>Used to indicate omissions in the text and recording. Also used when names referred to in the recording are withheld.</i>	reason	inaudible inaudible due to noise from audience - includes coughs, etc inaudible due to equipment difficulties inaudible due to hand over mouth inaudible, multiple speakers inaudible due to overlap ethics e.g. "i hope this will be removed" name ...	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
	extent	N words N secs uncertain	

2.11 <unclear>

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

2.12 <reading>

Element	Attributes	Values	Examples
<reading> <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"/> <reading> 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 </reading> <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.
-