

SPELLING CONVENTIONS

Adapted, with the authors' permission, from VOICE (Version 2.1 June 2007)

VOICE. 2013. *The Vienna-Oxford International Corpus of English* (version POS XML 2.0).

Director: Barbara Seidlhofer; Researchers: Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka, Nora Dorn.

1. CHARACTERS

a b c d e f g h i j k l m n o p q r s t u v w x y z	Only alphabetic Roman characters are used in the transcripts. No diacritics (e.g. ñ, ç or í) are permitted in the running text, even though they may be obligatory in the source language.
---	---

2. DECAPITALIZATION

<u>Example:</u> PP: phi:lip the se- philip the second's attempts to: collect art or	No capital letters are used in the running text except for marking emphasis (see mark-up conventions).
--	--

3. BRITISH SPELLING

British spelling	British English spelling is used to represent the speech of the participants. The second edition (1989) of the Oxford English Dictionary (OED) is used as the primary reference.
------------------	---

4. NON-ENGLISH WORDS

<u>Example:</u> AC: if there is if or there isn't a (.) latin america (.) identity not just colombia nor cuba hh (.) so (.) <L1sp> el reino de este mundo </L1sp> (.) <clears thout> talks about that (.)	Non-English words are rendered in the standard variant of the original language. The Roman alphabet is always used, with no diacritics.
--	---

5. FULL REPRESENTATION OF WORDS

<u>Example:</u> CP: yes so @@@ hh okay and then em yeah (.) e the <spel> a p a</spel> JW: yes CP: in test reference JW: yes	Although words may be pronounced with a foreign accent, they are generally represented in standard orthographic form. Only striking variations are noted (see mark-up conventions). If the <i>intended</i> form
--	---

CP: is the surname?	of a word is different from the word actually produced, the word actually produced is represented in standard orthography. In the example, the word uttered was test although the intended word was text .
---------------------	--

6. LEXICALIZED REDUCED FORMS

'cause	The only lexicalized phonological reduction that is reproduced throughout the transcript is that on the left.
gonna	The phonological reduction of 'going to' as 'gonna' is only reproduced when the speaker produces 'gonna to'.

7. CONTRACTIONS

i'm, they're, he's, we can't, he'd, etc.	Whenever they are uttered, all standard contractions are rendered.
--	--

8. HYPHENS

Example: DC: OR like you're saying we might have text-based questions as well (.) but i wouldn't normally give an excerpt from a (.) a novel	Hyphens are used according to British English spelling rules. The OED is used as a primary source of reference.
---	---

9. ACRONYMS

Example: EV: a question about nato don't just say everything you s- know about nato	Acronyms (i.e. abbreviations spoken as one word) are transcribed like words. They are not highlighted in any way.
Example: NT: um i'll see what i can do because obviously i know (.) a bit about the: telling ¹ exam but (Footnote 1: Topics in English Language and Literature)	Acronyms that are not current in everyday English are accompanied by an explanatory footnote.

10. DISCOURSE MARKERS

	All discourse markers are represented in orthography as shown below. The lists provided are closed lists.
yes, yeah	Backchannels and positive minimal

okay	feedback
mhm, hm, nn	(closed sound-acknowledgement token)
aha, uhu	(open sound-acknowledgement token)
oh, ah I see	(comprehension token)
no n-n, uh-uh	Negative minimal feedback
er, erm, um e em	Hesitation/filler
huh	tag-question
	Exclamations
yay, yippee, whoohoo, mm:	joy/enthusiasm
haeh	questioning/doubt/disbelief
a:h, ooh, wow, poah	astonishment/surprise
oops	apology
ooph, ff	exhaustion/frustration
ts, pf	disregard/dismissal/contempt
ouch, ow, ay	pain
sh, psh	requesting silence
oh-oh, u:h	anticipating trouble
ur	disapproval/disgust
oow	pity/disappointment