Wellington Archive of New Zealand English Transcriber's Manual

BERNADETTE VINE, GARY JOHNSON, JENNIFER O'BRIEN AND SHELLEY ROBERTSON

School of Linguistics and Applied Language Studies

Victoria University of Wellington

Language in the Workplace Occasional Papers

Number 5 (December 2002)

This series of occasional papers is aimed at providing a wide range of information about the way language is used in the New Zealand workplace. The first paper outlines the aims and scope of the core project, the Wellington Language in the Workplace Project, and describes the approach adopted by the project team in collecting and analysing workplace data. The second describes the methodology adopted to collect workplace interaction, and its developments and adaptations to the very different demands of disparate workplaces. Subsequent papers provide more detailed analyses of particular aspects of workplace interaction.

These include

- an analysis of varied ways people get things done at work, or the forms which directives take in different New Zealand workplaces
- an exploration of the functions of humour in workplace interaction
- an analysis of the structure of formal meetings in relation to the way decisions are reached
- a discussion of the concept of the "gendered" workplace and its relevance in New Zealand workplaces
- an examination of the varied literature on the role of e-mail at work
- an analysis of changes in some features of sexist language in workplace documents

The series is available in full text at this website: **http://www.vuw.ac.nz/lals/lwp**

Contents

Preface

The Wellington Archive of New Zealand English contains a number of corpora. The transcription conventions in this manual are based on the conventions used by the Wellington Language in the Workplace Project. These were adapted by the current corpus manager from the conventions developed to transcribe the Wellington Corpus of Spoken New Zealand English (WSC) and the New Zealand Component of the International Corpus of English (ICE-NZ).

The basic principles of this transcription system were established by the Corpus Research Advisory Group which consisted of Laurie Bauer, Allan Bell, David Britain, Janet Holmes, Graeme Kennedy, Chris Lane, Miriam Meyerhoff and Maria Stubbe. These were partly based on the conventions used by Chris Lane in his teaching and research, which in turn were based on conventions widely used in research at the time. Specific features were taken from Jefferson's CA conventions, Crystal and Davy (1975) and the work of Gillian Brown (Brown 1977; Brown, Currie and Kenworthy 1980).

The corpus system was refined as the transcribers encountered and met the obstacles of transcribing real data. Valuable input was contributed by the following people. Corpus Managers: Miriam Meyerhoff, Maria Stubbe, Raewyn Whyte, Sue Petris, Jane Pilkington, Jennifer O'Brien, Gary Johnson, Bernadette Vine; Transcribers: Alexander Tripp, Angela Lavender, Anissa Bain, Anita Easton, Ben Taylor, Bernadette Vine, Camille Plimmer, Claire Solon, Elizabeth Smith, Esther Griffiths, Gary Johnson, Jane Pilkington, Jen Hay, Jennifer O'Brien, Jenny Allan, Kate Kilkenny, Kate Wadsworth, Kerry McCarty, Lynette Sollitt-Morris, Margaret Cain, Martin Paviour-Smith, Meg Sloane, Michaela Stirling, Nina Flinkenberg, Penny Wilson, Rachel Lum, Rowena Samaraweera, Sarah Dreyer, Shelley Robertson and Sue Petris.

Many students and researchers have made inquiries about our transcription system. The publication of this manual makes our conventions readily accessible.

Bernadette Vine

Corpus Manager

Wellington 2002

Main transcription conventions

All examples used in this manual are either fabricated or adapted from WSC extracts.

Pseudonyms

Pseudonyms are used to label speakers and people mentioned. This is generally a name with the same gender or ambiguity of gender (e.g. sue-->jill, chris(tine/topher)--> pat(ricia/rick), stress patterns, number of syllables and ethnicity (e.g. tama-->hemi). The assigned initials of speakers are those of the assigned names so that, for example, someone assigned the pseudonym Fred Smith will be identified as FS throughout the transcript.

Times

The time at which the extract begins on the tape is noted at the beginning of the body of a transcript. Every whole minute is noted in the margin, e.g.,

2:54	[side one]
HD:	okay so we need to arrange a time for a team meeting have you got a
3:00	copy of that memo i sent out i seem to have forgotten my copy'll just
	get it and i'll get you a copy too
3:15	
	[one minute silence as HD goes to get the memo]
4:15	
TS:	did you find it okay cos i'll have mine somewhere
HD:	yeah no here it is so +++ i'd suggested people keep thurs and fri
	afternoons free but clive has emailed me to say he has that course on

	friday afternoons will thursday be all right with everyone in your section
TS:	yep kaye is also doing that course so thurs will suit her and noone else
	indicated that they had any clashes
HD:	great we'll say thursday then at two
TS:	right +
HD:	thanks will you let your group know and ĩll email my lot
5:00	
TS:	sure see you then
5:02	
	[end of interaction]

Standard character set

The permissible characters used in transcription are:

---

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

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

( ) [ ] : - # + = / \

- Alphabetic Roman characters are used in lexical transcription and editorial comments.

- No diacritics or non-roman characters are permitted.

- Upper case is reserved for marking emphatic stress.

- Non-alphabetic characters are used to mark discourse features, editorial comments and their scope.

- Parentheses enclose doubtful transcription, square brackets enclose editorial comment and the colon indicates its scope.

- The hyphen indicates an incomplete word.

- The hash indicates an otherwise ambiguous clause boundary.

- The plus sign shows a pause.

· The slash and equals sign are used to show simultaneous or overlapping speech

These features are outlined below with examples.

Punctuation

No punctuation is used (except for apostrophes). Capitals are used only for emphatic stress.

Stress

CRAZY, **UNbelievable**   Capitals are used to indicate emphatic stress.

Comprehension problems/transcriber doubt

( )	Untranscribable or incomprehensible speech
(well)	Transcriber's best guess at unclear speech
??:	Unknown speaker
AT?:	Unknown speaker, possibly AT
#	To signal end of "sentence" where it is ambiguous on paper
?	Use only to signal "question" where it is ambiguous on paper:
	e.g. are you going to the zoo tomorrow
	e.g. you're going to the zoo tomorrow?

Pauses

+	Short pause of up to one second
++	One to two second pause
+++	Two to three second pause
(4)	Four second pause: i.e. after three seconds indicate length by noting number of seco parentheses. Noting the approximate length of the pause, especially if it is more than mainly done as a guide for the listener.

Noises

er	all hesitations not ending in -m
um	all hesitations ending in -m

mm	minimal feedback
mhm	yes
uh uh	no
aha	yes
[tut]	bilabial/alveolar/dental clicks
[voc]	funny untranscribable noises not covered by any other convention. IPA is used when

Incomplete words

wha-	hyphen indicates cut off word, both self-interruption and other speaker interruption
AA:	and then he ca- came to see me
but not here,	
BB:	have you seen oh while i remember i must tell you something

Hyphens are not used where a word is repeated in its entirety e.g.

CC:	and after that they they went to the movies

NB: when a word normally contains a hyphen in standard spelling, e.g., by-product, it is written as one word without the hyphen, i.e., byproduct.

Simultaneous speech and continuous utterances

//	Indicates start of simultaneous or overlapping speech in utterance of "current" or "fi
\	Indicates end of simultaneous or overlapping speech in utterance of "current" or "fir
/	Indicates start of simultaneous or overlapping speech in utterance of "incoming" or
\\	Indicates end of simultaneous or overlapping speech in utterance of "incoming" or " e.g.
	AA: i'd like to come as well + //is\ that okay
	BB: /yeah\\

Overlaps are marked on pauses, as in the following example. This means that each utterance is kept together e.g.

	AA:	i'd like //+\ to come as well
	BB:	/mm\\

Rather than e.g.:

	AA:	i'd like
	AB:	mm
	AA:	to come as well

Numbering of Overlaps

Numbering is added where a speaker is overlapped more than once within a turn:

e.g.	AB:	you've got to deep fry them 1//do\1 you + or just
		2//pan fry them\2
	CG:	1/mm\\1
	CG:	2/no + they were\\2 very greasy

Numbering - a further example

AK:	remember that time we went to gisborne 1//with\1 the martins
	and we went to that 2//place with the huge\2 rocks
BD:	1/yeah\\1
CL:	2/and it rained\\2
CL:	and it 1//rained\1 the whole time 2//and none of us\2
	had raincoats
AK:	1/yeah\\1
AK:	2/i'd forgotten that\\2

Although the two interruptions on AK are made by different people (first BD, then CL), they are still numbered consecutively as interruptions made within a single turn of AK's.

When the speaker changes (in this case from AK to CL), and a different person is being interrupted (in this case CL), the numbering of interruptions on CL starts from 1 again.

Paralinguistic and Relevant Non-verbal Features

Comment tags for paralinguistic and relevant non-verbal features e.g. [quickly], [drawls] come BEFORE the relevant section of speech, which is identified using a colon at the beginning and end of that section to which the comment tag applies. (NB notes on non standard pronunciations come AFTER the word, see note below).

e.g.:	er and all the other stuff + but [quietly]: we won't talk about that: SO how many sh milked lately
e.g.:	william gunn records quote [reads]: caused us great alarm: now the alarm was felt

These tags (and indeed all comments) appear in SQUARE BRACKETS. Annotations use capitals where appropriate (proper nouns, etc) and follow standard spelling. Comments are in adverb, 3rd person singular, or "in/with a _____" form e.g. [quietly], [whispers], [with a fake American accent], [in Maori].

Tags we have used include:

[quietly]	for quietly, softly	
[drawls]	for slowly - to signal drawn out words	
[exhales]	for audible exhaling	
[inhales]	for audible inhaling	
[sighs]	for voiced exhaling	
[laughs]	tags like this one can occur either independently or over an utterance	
	e.g. "[laughs]: oh i can be like that too: [laughs]"	
[quickly]	[whispers]	[coughs]
[shouts]	[sings]	[groans]
[snorts]	[sniffs]	[clears throat]
[reads]	[paraphrases]	[with silly voice]
[with fake American accent]		

Note that the end colon only ever appears at the end of a word, even if the tag does not apply to the whole word (i.e. "[laughs]: yesterday:" and NOT "[laughs]: yester:day").

Group Speaker Identification of Simultaneous Laughter

All:	If all speakers laugh at once and there are *more* than two speakers in the extract, the is labelled in the margin as All:,
	e.g.: All: [laugh]
Both:	If two speakers laugh at once and there are *only* two speakers in the extract, then lat in the margin as Both:,
	e.g.: Both: [laugh]
[others laugh]	If two or more speakers' laughter overlaps a third speaker's speech then this is trans laugh] within the simultaneous speech slashes
	e.g.: AG:	jeremy said //no way will\ i do that
		/[others laugh]\\

If only some of the speakers are laughing simultaneously with another's speech while the other participants remain silent then this is transcribed by noting which speakers laugh in square brackets within the simultaneous speech slashes, e.g. (in this example the other speaker, DJ, is silent):

AG:	jeremy said //no way will\ i do that
	/[BB and CP laugh]\\

Non-standard Speech

General principle: Non-standard speech is transcribed in the standard orthographic form closest to the full morpheme so that it can be picked up in word frequency counts. The only exceptions to this are very frequent variants with familiar (standardised) variant forms, e.g. *cos*, *gonna* but *he* not *'e* and *stamping* not *stampin'* (see below).

A more comprehensive list of spelling decisions can be found in Holmes, Vine and Johnson (1998).

yes no	)	
yeah nah	)	can be used to transcribe variants of *yes/no*.
yep nope	)	
okay		is the standard spelling (i.e. NOT *ok* or *OK*)
okey doke(y)		rhyming form of *okay*
and		represents all variants, e.g., *and, 'n, 'nd* etc
er		represents any vocalised hesitation except *um*. Any noteworthy idiosy particular speaker can be recorded on the cover sheet if appropriate.

ah		represents the vocalisation in expressions like *ah* i see what you mea NOT a hesitation marker. The vowel may be relatively long or short transcribed as *ah* regardless.
eh		tag, e.g. "badjelly is really cute *eh*".
oh		represents all utterances in the *oh - ooh* group
cos		represents all abbreviated variants of *because*
gonna		pronunciation of *going to*. However all other non-standard verb form transcribed in full (e.g. "hafta" as *have to*, "wanna" as *want to*, etc)
should've		either *should have* or *should've* unless *should of* is said with a distinc this case it is marked "should of [pronunciation of have]".
fella(s)		with final schwa represents colloquial use of *fellow(s)*
blimmin		colloquial use of blooming, as in *blimmin heck*
me		*me* in the example "put it in *me* mouth" transcribed as *me* and marked [pronunciation of my]".
jeez		for the contracted form of *jesus*
righty oh		based on *right oh*
about, them		*'bout* for *about* and *'em* for *them* are considered standard pronunciati transcribed *about* and *them* respectively, without annotation.
whatsitsname	)	
whatshername	)	for person/thing whose name is unknown,
whatshisname	)	forgotten, or deliberately overlooked
whatsit	)	
whoohoo		expression of enthusiastic excitement
anti smoking	)	transcribe *anti-* compounds with space between
anti drugs	)	elements
gaw		for God pronounced /go/
gawd		for God pronounced /god/
thingumabob	)	for person/thing whose name is unknown,

thingumajig	)	forgotten, or deliberately overlooked
thingummy	)	
uh huh		*uh-huh*, expression of agreement or acknowledgement

Contractions

'll, 've, 'd, 'nt etc are used as appropriate, regardless of whether the clitic is attached to a verb or noun host. For example "the fellas*'ve* done it before". Auxiliary clitics can also appear outside negative clitics, e.g. "she mus*tn't've* been conscious at the time". (See also *cos, gonna, should've, blimmin, jeez* above.)

Non Standard Pronunciations

koa for *kea, worfore* for *warfare*, plus malapropisms, spoonerisms, etc are transcribed as the orthographic standard, with a comment in square brackets where appropriate.

e.g.   ....talking about trench warfare [pronounced as warfore]...

e.g.   ...and one of my nephews [pronounced /nebjuz/]...

NB: *brought* pronounced as *bought* is marked like this:

...brought [pronounced as bought]...

Where a pronunciation is widespread, e.g. *libarian* for *librarian*, the orthograhic standard is used and annotations are not added.

Numbers, Acronyms, Abbreviations and Contractions

DOC (Department of Conservation) can be said as either doc /dok/ or d o c /di ou si/. It has been transcribed as doc or d o c respectively. Other examples: anzus, anzac, v d, r s a, m ps

Numbers and forms that are usually abbreviated are written out in full,

e.g. "nineteen oh three" for *1903*

"et cetera" for *etc*

"saint" for *St*

"okay" for *o.k.*

"missus" for *Mrs*

"mister" for *Mr*

"miss" for *Miss*

"ms" for *Ms*

Utterances in Languages Other Than English

The comments in the following sections apply to the WSC and ICE-NZ.

In transcribing the WSC and ICE-NZ it was decided that where a transcript contains words/phrases that are not considered part of New Zealand English, it would be useful to mark this in square brackets before the word/phrase.

e.g. [in German]: wunderbar: [in Maori]: kia ora:

Maori

Although many Maori words are part of NZE, overseas researchers are likely to be unfamiliar with them. Therefore Maori words and phrases, with few exceptions (see below), are marked as [in Maori] and glossed on the cover sheet. ANY occurrence of a Maori word (except proper names, e.g.*Paraparaumu, Mrs Ranapia*) were glossed on the cover sheet.

Long Sections in Maori

Long sections in Maori, or any other language, (i.e. extra-sentential code switching) are not transcribed. The length of the section is noted in square brackets and a brief summary is included where possible e.g.

> [Two minutes thirty two seconds in Maori - AD and BL discuss AD's father's recent ill-health and operation]

Maori Words not Marked as [in Maori]

Pakeha, *Maori*, *Aotearoa* and *marae* were the only words from Maori which are not marked as [in Maori]. Four exceptions to this principle were:

1.   If one of these four words is used with a different meaning than the usual NZE sense, e.g. if *Pakeha* is used in the sense "English language", then it is considered to be a Maori word and transcribed "[in Maori]: pakeha:".

2.   If the word is used in a Maori phrase. e.g.*Maori* in *wairua maori* would be transcribed "[in Maori]: wairua maori:"

3.   Maori names for native flora and fauna are*not* marked as [in Maori] when either (i) there is no English alternative, or (ii) the Maori name is more frequently used than the English equivalent. All Maori words were still glossed on the cover sheet, e.g. (i)*paua* unmarked but glossed, (ii)*tui* unmarked but glossed (English equivalent the rare "parson bird"), (i/ii)*kauri* unmarked but glossed.

4.   Maori tribal names, e.g. "ngati whatua" for*Ngaati Whatua*.

Proper Nouns

Proper nouns are not marked [in Maori], e.g. for the placename*Paraparaumu* it is not noted whether the Maori or the anglicised pronunciation (or some variation on either) is used.

Plurals/Clitics/Inflectional Endings

This note applies to any language in which English morphemes are added to non-English bases, but is

most relevant to Maori. If a clitic (e.g. possessive *'s*) or an inflectional or plural ending is added to a non-English word (e.g. *a couple of hangis*) then the whole orthographic word is enclosed within the colons, e.g. "a couple of [in Maori]: tangis:". A note was then made on the cover sheet "**tangis** - plural of **tangi**, a funeral".

Guidelines for Deciding Whether or Not a Word is Part of NZE

It is not always easy to decide whether a word is part of NZE or not. It depends partly on whether the phonology of the other language is adopted, or whether the word is anglicised to some extent. The following guidelines were developed to help transcribers reach a decision:

1. Do you understand the meaning of the word/phrase? Do fellow Corpus workers?

2. Is the word/phrase in an English dictionary, especially a NZE dictionary e.g. Heinemann's?

If in doubt the transcribers were advised to err on the side of caution and mark the word on the principle that too much information is better than too little.

Where a transcript contains a number of words in a language other than English, and it is useful for the listener to know the meaning of these words, they were glossed on the cover sheet, e.g.,

Maori words used in this transcript

haka - dance, song accompanying a dance

moko - short for *mokopuna*, i.e. Grandchild

marae - Maori meeting ground

Cases where this may be necessary include where one or more speaker is Maori or in tutorials where foreign terms are used.

Example of a transcript

What follows is a modified extract from WSC transcript DPC015.

===

[side one]	
12:43	
DW:	after i saw you yesterday i decided to call it a day we //went-\
MM:	/well\\ that's wise
DW:	we went 1//up to\1 um this disused railway line up in um +

	i 2//don't (know)\2
PP:	1/mm\\1
MM:	2/rimutakas\2
JW:	2/it wasn't\\2 rimutak- //temarua\
DW:	/was it rimu\\takas
JW:	//temarua ( )\
MM:	/yeah yeah\\
DW:	and we walked through a tunnel there and it was really scary so
	edward started being very distressed in the middle as you say it's it's
13:00	1//y- you (if you get between them) it was entirely 2//dark\1\2
JW:	1/in the middle- like it was all right at each end\\1
AM:	2/yeah ( )\\2 [laughs]
JW:	(towards) in the middle it was absolutely dark and edward just shut up he wasn't gor say anything

References

Brown, Gillian 1977. *Listening to Spoken English* London: Longman

Brown, Gillian, Karen L Currie and Joanne Kenworthy 1980. *Questions of Intonation* London: Croom Helm.

Crystal, David and Derek Davy 1975. *Advanced Conversational English* London: Longman.

Holmes, Janet, Bernadette Vine and Gary Johnson 1998. *Guide to the Wellington Corpus of Spoken New Zealand English*. Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.

THE ARCHIVE OF NEW ZEALAND ENGLISH: Linguists at Victoria University of Wellington have been involved in the collection of New Zealand English for three different corpora, one spoken, one written, and a third which includes both spoken and written data. The transcripts from the corpora are now available as text files on CD.http://www.vuw.ac.nz/lals/corpora.htm#The Wellington Corpus of Spoken New Zea

LANGUAGE IN THE WORKPLACE