Open Research Online

The Open University's repository of research publications
and other research outputs

"Do I speak better?" A longitudinal study of lexical chunking in the spoken language of two Japanese students

Journal Item

For guidance on citations see FAQs.

Version: Version of Record

Link(s) to article on publisher's website:
http://www.brookes.ac.uk/schools/education/eal/eal-2-2/vol2-no2-leedham.html

oro.open.ac.uk

"Do I speak better?" A longitudinal study of lexical chunking in the spoken language of two Japanese students

Maria Leedham,
The Language Centre, Oxford University

Abstract

The prominence of lexical chunks or prefabricated language has grown over recent years, however there have been few longitudinal case studies exploring changes in non-native speaker (NNS) speech and little work done involving NNSs in identifying chunks in their own speech. This study attempts to track changes in two intermediate-level Japanese students' spoken usage of lexical chunks over a period of five months in the UK. Each NNS was recorded three times in conversational long turns at two-month intervals.

Twelve native speakers (NSs) were asked to order transcripts of each student's speech by perceived fluency level and three also underlined the lexical chunks; however there was little coherence amongst NSs in these tasks. Identification of chunks using Wordsmith software suggests an overall rise in the percentage of talk within chunks and a reduction in ill-formed chunks over the five months.

Following some awareness-raising training on identifying lexical chunks, the Japanese students themselves were asked to identify chunks within their own transcripts. Despite the difficulty of the task, they were able to do this and additionally offered insights into which chunks were common for them. These insights included an awareness of typical Japanese phrases and how they felt their speech had changed overall. A further recording and transcribing cycle suggests that this training resulted in some short-term uptake as the percentage of chunks used increased after the lessons. Both students found it highly motivating to record and analyse transcripts of their talk as they could see progress in their own spoken language development.

Introduction

"*I suspect that we speak mostly by stitching together swatches of text that we have heard before.*" Becker, 1975:3081.

Becker's memorable description encapsulates a now widely-accepted view of language (e.g., Lewis, 1993; McCarthy, 2001). However the identification of chunks is not straightforward as, while computer software such as Wordsmith Tools can list "clusters" or phrases which recur, it cannot reliably identify chunks which may occur only once in a million-word corpus (Moon, 1998:72). On the other hand while native speakers (NSs) are able to intuit the chunks in a text, one person's intuition may not be the same as another's (Foster, 2001). It is especially difficult for NSs to pick out chunks used by non-native speakers (NNSs) since a text may contain phrases which the NS considers ill-formed but which are chunks within a particular NNS's lexicon.

Foster (2001) recorded and transcribed NNS and NS talk and examined the percentage of their language that was produced within chunks. Her research suggests that NNSs use fewer lexical chunks in their speech than NSs and also use a less varied range of chunks.

Where this study differs from Foster's is in considering two students' progress in oral fluency over a five-month period. Twelve NSs were first asked to read three transcripts from each of two intermediate-level Japanese NNSs and to decide on an order of fluency. Following this, three NSs and this teacher-researcher chunked the transcripts.

A second research cycle looks at how awareness-raising tasks may affect chunking and suggests this is both possible and desirable. This has significant pedagogical implications although research on a larger scale is needed.

Research questions

The study began with the following research questions:-

1.   Does NS ordering of the transcripts according to perceptions of fluency correspond with the chronological order?

2.   Can NS linguists reliably identify lexical chunks in NNSs' transcripts?

3.   Can NNSs identify chunks in their own transcripts?

4.   Does NNSs' talk consistently increase in both number and variety of chunks over time?

The first research cycle set out to answer these questions and consideration of the findings led to the development of a second cycle.

Chunks and fluency

The term "lexical chunks", also known as "memorised sentences and lexicalised stems" (Pawley and Syder, 1983), "micro-units" (Foster, Tonkyn and Wrigglesworth, 2000) or "formulaic expressions" (Norton, 2001) covers items from compound nouns like "mother tongue" to longer lexicalised sentence stems such as "it's really hard to understand". While each researcher's terminology and range may be slightly different, all broadly adhere to Sinclair's "idiom principle" of language building up "semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (1991:110).

Foster, Tonkyn and Wigglesworth (2000:355) state that "the more proficient speaker will be the person who can keep track of more complex micro-units", that is, who can quickly access multiple chunks when speaking. This implies that learning language in chunks not only renders the speaker more "native-like" (Pawley and Syder, 1983) and thus more accepted by a native speaker speech community, but also helps the learner to speak more fluently since longer stretches of discourse can be readily produced.

One increasingly common measure of fluency in applied linguistics research is to calculate stretches of language produced as wholes or 'chunks' within lengthy turns and probably produced without pauses. Other studies have measured disfluency by counting filled and unfilled pauses,

false starts, hesitation devices and self-corrections (see Foster, Tonkyn and Wigglesworth, 2000 for a full discussion).

This study uses lexical chunks as a measure of fluency in speech and considers the extent to which an increase in quantity and quality over time indicates an improvement in NNS speech. Foster's (2001) method of calculating words within chunks as a percentage of the total number of words is followed. Quality is discussed in terms of the percentage of chunks which are repeated or are in some way ill-formed.

Transcribing conversational long turns

This study concerns the language produced by two Japanese females of similar age and background within one-to-one tutorials over a five month period. The students (known as 'M' and 'C') were recorded three times at two-month intervals. On each occasion after a preliminary chat the student was asked to think of an incident in the previous week in which they had had difficulty in either communicating or in general everyday life. This type of conversational long turn is familiar to the students within a tutorial setting and produced different topics on each occasion.

The six recordings were then transcribed in order to eliminate aural cues from the research since this study is not concerned with pronunciation. For each recording the NNSs' first recorded long turn was transcribed until its natural end-point, resulting in around 250 words each time (see appendix 3 for extracts of M's transcripts).

NS ordering of NNS transcripts

The transcripts were given to six NS linguists (five English as a Foreign Language (EFL) teachers and one computational linguist) and six NS non-linguists (four engineers, one school secretary and one information analyst) who were separately asked to order them for each speaker according to the level of fluency. That is, each of the NSs were asked to read three transcripts per NNS and to decide which of the three was least, next and most fluent. It was expected that there would be some correlation between this ordering according to fluency and the chronological order in which the long turns were elicited, assuming that the NNSs made progress over time and that this progress was revealed in the transcripts of their talk.

However, there was little consistency between NSs and no overall relationship between a tally of the orderings and the chronological order of the transcripts. Just three out of the twelve could determine the actual order for student M's transcripts and only one NS for those of student C (see appendix 1).

No guidance was given as to what to look for in ordering the transcripts since this left the question open as to which factors contribute to greater "fluency" in the view of the NSs. Interviews following the ordering exercise revealed that the NSs were looking for measures of 'improvement' rather than increased fluency, that is, they noticed an absence of errors and an increase in 'difficult' lexis. For some, the number of filled pauses was a prominent mark of disfluency while for others grammatical inaccuracies were most salient.

Comments could contradict each other, thus for one NS M's second transcript was "more fluent" while for another it was "very broken". The NSs used content 'clues' wherever possible giving comments such as that a particular talk must be the most recent because the student "seems to be more confident when things go wrong" or that the student was "finding her feet".

NSs chunking NNS transcripts

The next stage was to ask three of the NS linguists to chunk the NNS transcripts. Following a brief practice using a NS transcript, the three transcripts for each student were returned in random order and NSs were asked to underline any words which they felt were produced as one item rather than built up structurally (see appendix 2). Interestingly, the informants sometimes revised their earlier ordering after studying the transcripts and underlining the chunks (see appendix 1).

Two NSs found very few chunks within the transcripts and each found chunks which the other two did not classify. The range was large; taking C's 'housemates' transcript as an example, two NSs underlined 15% and 18% of the text respectively as composed of chunks while the third (RP) chunked 30%. This is partly accounted for by the tendency of one NS (RP) to underline longer stretches of text; for example where another NS underlined "mother tongue", RP chose the longer "it's not my mother tongue". RP also picked out chunks which neither of the other NSs identified, such as "it's difficult to understand directly", "easy to forget" and "every night". Frequently groups of words were identified by one NS as chunks but ignored by the other two; examples are "the first time", "every weekend", "that kind of things" and "I can't understand".

The most notable overall finding is that chunking did not appear to increase over time. It is unclear why the percentages in the second talk are the highest but this may be due to facets of those talks rather than a change in fluency patterns. Additionally, the chunking pattern does not correspond to the ordering of transcripts by NSs.

It seems, then, that there is little agreement as to what constitutes a chunk or where a chunk begins and ends, and that determining chunks is especially difficult when analysing NNS speech. In particular there are many 'missed chunks' and discrepancies between the three NSs' analysis.

NNSs chunking their own transcripts

The NNSs were also asked to chunk their own transcripts after some practice in determining chunks. Despite the difficulty of the task (J. Willis, 2005, personal communication) both could accomplish this and in fact M classified twice as much of her third transcript as within chunks as did the group of NSs (app.3). While some of these such as "I was **so** embarrassed" are repeated, others such as "of course" and "you know" were simply missed by the NSs.

M commented that the chunk "it will takes about one week" from the first transcript was incorrect but said she had regarded this as a chunk at the time of the recording. Similarly, C recognised that "quite easier" was a chunk for her but was incorrect. This suggests that underlining chunks is a useful learning activity to encourage students to notice their own common errors.

Using Wordsmith to chunk

It seems that there is little correlation between fluency as perceived by all twelve NSs and the percentage of the text in chunks as perceived by three NSs; moreover the whole procedure appears haphazard as so many chunks were missed by the NSs.

At this point I decided to chunk the transcripts myself. Other researchers such as Baigent (2005) and Nesselhauf (2003) have also resorted to relying largely on their own intuitive knowledge of

chunks. To reduce the chance of overlooking phrases which are chunks for the NNS yet not salient to a NS, Wordsmith's clustering feature was used.

For Mike Scott, the developer of Wordsmith Tools, "all words have a tendency to cluster together with some others" (Wordsmith version 4.0, help pages) These "clusters" differ from chunks in that the latter can be "identified intuitively by people" and "feel 'whole' in some way" (Willis, 1998) while clusters can be words from different clauses which frequently co-occur such as "go but I". Thus software such as Wordsmith can pick out n-grams occurring twice or more within processed text(s) but cannot determine if these are psychologically noticeable as chunks.

The transcripts were chunked twice, with a three-week interlude between. The number of chunks found rose in every case on the second occasion, largely due to a greater acceptance of the phrases that can be considered a chunk. What does seem to consistently be the case is that the more time spent, the greater the number of identified chunks. For example, perhaps "I went to the bookshop" is so frequently used by a student that it becomes a chunk for them.

My own chunking suggests that the chunks did not consistently increase in quantity or quality; so for example there was no move towards longer lexicalised sentence stems in the later talks. In M's case when repeated and ill-formed chunks are removed, the percentage of words within chunks went up over time using the first calculations, but not using the second. In C's case both sets of figures show an increase in chunks from the first to the second recordings then dropped dramatically in the third.

Percentage of transcript within chunks – student M

	underground		food		washing	
X: % of text in chunks	*35*	**48**	*43*	**52**	*39*	**39**
Y: % in repeated chunks	*2*	**4**	*4*	**4**	*2*	**3**
Z: % in ill-formed chunks	*12*	**12**	*8*	**13**	*4*	**5**
X – (Y+Z) ie % of text in chunks minus repeated and ill-formed chunks	*21*	**32**	*31*	**35**	*33*	**31**

Key: *1* = percentage found on first occasion of chunking
1 = percentage found on second occasion of chunking
NB: Where 'repeated' and 'ill-formed' chunks overlapped, the chunk was included in the 'ill-formed' % only.

Percentage of transcript within chunks – student C

	Bike & bars		housemates		movies	
X: % of text in chunks	*52*	**62**	*46*	**56**	*31*	**35**
Y: % in repeated chunks	*9*	**6**	*7*	**5**	*5*	**5**
Z: % in ill-formed chunks	*15*	**24**	*7*	**6**	*4*	**9**
X- (Y+Z) ie % of text in chunks minus repeated and ill-formed chunks	*28*	**32**	*32*	**45**	*22*	**21**

What is clear is the dearth of long chunks in the NNS transcripts. Utterances are built up in two and three word chunks with pauses and hesitations separating them and there are very few of Pawley and Syder's "memorised sentences and lexicalised stems" (1983).

Ill-formed chunks and "Just the Word"

Significant in this analysis are chunks which were incorrect or unusual in some way and which seem 'ill-formed' by NS norms due to mis-collocations or grammatical errors. These ill-formed chunks are classified as chunks for the students because they are repeated by the NNS within the three transcripts and thus identified by WordSmith Tools as clusters.

Ill-formed chunks fell into two categories: grammatical and collocational. The former include mistakes with articles ("have a dinner"), misuse of the passive ("stolen your bag") and comparative adjective formation ("it's quite easier"). Others are mis-collocations such as "I noticed about it" or "change my taste". The trickiest decisions over whether a chunk is ill-formed occurred with collocations which could be blocked or could simply be weak. For example, the following chunks taken from the transcripts are common in Japanese students' speech but seemed odd in some way:

> "during this term"
> "exchange opinion"
> "we Japanese women"
> "around the weekend"

At this point Just The Word software was used to determine the strength of these collocations. This uses 80 million words of the BNC to determine the strength and frequency of collocations. There were no instances of "during this term", "exchange opinion" and only two occurrences of "around the weekend" thus these chunks are classed as blocked collocations by Just The Word.

Some blocked collocations seemed to be translations of Japanese phrases. For example "the taste is good" is a word-by-word translation of the Japanese equivalent. Baigent (2005) and Nesselhauf (2003) similarly found that students' errors were strongly influenced by their L1. The two NNSs were able to identify some of these errors and realise that they were not standard chunks.

For both NNSs, the percentage of words within ill-formed chunks goes down over time from first to third transcripts, as would be expected in the case of full-time students of English living in the UK.

Variables

Since NSs did not appear to be judging the fluency of the talks by an increase in the number of well-formed chunks, what else was going on? One factor is that the two NNSs became more confident with the one-to-one tutorial arrangement and the recorded tutorial long turn. This perhaps increases risk-taking with language, resulting in a search for more creative or original language use and may result in more pausing and more errors. While attempts were made to elicit 'new' topics from students by asking them to describe a recent incident, it is impossible to ensure that the topics chosen were unrehearsed.

Both students in this study are Japanese and may have been inclined to repeat chunks since Japanese allows repetition of the same lexical item whereas English favours a greater use of synonyms (Swan and Smith, 1987). White suggests that Japanese has more clause boundaries than English and thus more opportunities for listeners to provide backchannelling to show 'omoiyari' or

harmony (1989). Pausing from the subjects could thus be viewed as seeking confirmation that they still hold the floor, yet in a transcript it may be seen as a measure of disfluency.

An action research cycle: recording, training and re-recording

After the two students had chunked their three recordings a second research cycle was conducted to investigate the awareness raising effect of teaching about chunking. Within a class students were asked to recall the previous analysis activity on the NS transcript, then pairs of students were given a transparency film and asked to place this over the text. Using different coloured pens, one pair underlined 'grammatical chunks', one focused on 'sentence beginners' and the third individual student looked at 'vague language'. The three transparencies were then overlaid and additional chunks of each category type were elicited.

Two days after this lesson a fourth recording of each student was made and chunked using Wordsmith. Interestingly, the percentage of each student's text contained within chunks increased greatly in comparison with the previous three recordings. For M the previous maximum minus repeated and ill-formed chunks was 35% whereas the fourth recording had 42% of the text within chunks. For C the difference was more dramatic with a previous peak of 45% and a final rating of 57%. These figures are similar to Foster's findings of chunking rates in NS speech. Further research would be needed to determine whether this improvement is a short-term gain or part of a long term change and the many variables outlined above also have some part to play.

Notably, in discussion of the final recording M commented "always when I speak English I reassemble an English sentence like a puzzle". She also made reference to translating, saying "when I think of a Japanese sentence and I try to speak in English I…I have to say… ahh… last word in Japanese". The act of analysing her own previous utterances helped M to think about how she spoke in English.

Pedagogical Implications

For Willis and Willis, using transcripts "allows learners time to notice features that may not be noticed for a long time if only heard in the flow of real-time conversation" (1996:76). Willis (1998) discusses the use of written and spoken texts as part of a learner's "pedagogic corpus" which is built up throughout a course of study. Later awareness raising activities can make use of all the texts which have preceded the lesson. While corpus study may seem initially overwhelming for students, guidance by the teacher can make it a little more accessible. Learners need to realise that chunks can rarely be literally translated word for word and that translation is best carried out chunk by chunk (Swan in Schmitt and McCarthy 1997:127).

The second cycle of awareness-raising, recording and transcribing was particularly fruitful and has since formed the basis of several tutorial sessions with Japanese students. Students appreciate the recording, transcription and subsequent analysis of their own talk since this provides them with a way to measure their oral progress over time.

Conclusion and future research

This study set out to determine the extent to which lexical chunking can be used as a measure of NNSs' progress in oral fluency. While there does appear to be an overall upwards trend in the percentage of students' language within chunks over the time of the study and a corresponding

decrease in ill-formed chunks, this only roughly correlates with NSs' perception of improved fluency.

For the three NS linguists, the chunking task was frustrating and ultimately rather haphazard. Even when the texts were chunked with the assistance of Wordsmith, the two attempts differed considerably. It seems that the more time spent looking for chunks in a transcript, the more will be found. Even with guidelines giving fairly explicit definitions of types of chunk there are still many unclear cases.

Controlling the planning aspect of the long turns would allow greater confidence in the degree to which the recordings represent a learner's progress over time. This could be done by encouraging NNSs to talk about a new topic, that is, to explicitly ask them to describe an incident that they had not previously told anyone in either English or their first language.

The pedagogical action research cycle does appear to show that awareness-raising is a useful activity which may serve to increase the use of chunks; this is an interesting area worthy of replication. To what extent does such training help native and non-native speakers to recognise chunks in the target language or their native language? Can this help learners to use more, longer and more varied chunks? This small research cycle suggests that a little training has a great effect and it would be interesting to follow this up at a later date to see if the increase is sustained.

References

Baigent, M. (2005) 'Multi-word chunks in oral tasks'. In Willis, J. & C. Edwards (Eds.) *Teachers Exploring Tasks*. Palgrave Macmillan.

Becker, J. (1975) 'The Phrasal Lexicon'. Advanced Research Projects Agency report.

Foster, P, A.Tonkyn & G.Wigglesworth, (2000) 'Measuring spoken language: a unit for all reasons'. *Applied Linguistics* 21/3

Foster, P. (2001) 'Rules and Routines: a consideration of their role in the task-based language production of native and non-native speakers'. In Bygate M., Skehan P. and Swain M. (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing.* Longman. pp 75-93

Just The Word concordance software: http://www.sle.sharp.co.uk/JustTheWord/index.html

Lewis, M, (Ed.) (2000).*Teaching Collocation*, Language Teaching Publications

McCarthy, M. (2001) *Issues in Applied Linguistics.* Cambridge University Press

Moon, R. (1998) *Fixed Expressions and Idioms in English.* Clarendon Press Oxford

Nesselhauf, N. (2003) The use of collocations by advanced learners of English and some implications for teaching. In *Applied Linguistics 24/2: 223-242*

Norton, J. (2001) 'Formulaic expressions in English and Japanese: implications for teaching and learning'. In *The Japanese Learner: Context, Culture and Classroom Practice*, Gray, K., M-A. Ansell, S. Cardew & M. Leedham (Eds) Oxford University Department for Continuing Education. Available online at: http://www.brookes.ac.uk/schools/education/eal/jl-archive/jl-bestof.html!

Pawley, A. & F.Syder, (1983) Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J. & R.Schmidt, (eds.) 1983 *Language and Communication.* Longman

Sinclair, J. (1991) *Corpus, Concordance, Collocation* OUP

Swan, M. & B. Smith, (eds.) 1987 *Learner English. S*ection on Japanese speakers. CUP

White, S. (1989) Backchannels across cultures: a study of Americans and Japanese. In *Language in Society*. March 1989

Willis, J. &Willis, D. (Eds.) (1996) *Challenge and Change in Language Teaching.* Macmillan Heinemann

Willis, J. (1998) Taped lecture given on lexical chunks

Wordsmith Tools version 4.0, developed by Mike Scott and OUP, downloaded from www.lexically.net//wordsmith/purchasing.htm

Appendix One : NS perceptions of fluency by ordering of transcripts

Student M: Order

		underground (first)	food (second)	washing (third)
NS Linguists (by initials)	AP	3, **1**	1, **2**	2, **3**
	PW	1, **1**	2, **2**	3, **3**
	RP	1, **3**	3, **2**	2, **1**
	RG	1	3	2
	RLy	3	1	2
	SW	1	2	3
NS non-linguists	RL	1	3	2
	SM	2	1	3
	JL	1	2	3
	JW	2	3	1
	VL	1	3	2
	KH	2	3	1

NB: Numbers in bold show the changed order given by the three NS linguists after they had underlined chunks in the transcripts.

Student C: Order

		bike and bars (first)	housemates (second)	going to movies (third)
NS linguists	AP	2, **2**	3, **3**	1,**1**
	PW	3, **2**	2, **3**	1,**1**
	RP	3, **3**	2, **2**	1,**1**
	RG	3	1	2
	RLy	2	1	3
	SW	1	3	2
NS non-linguists	RL	3	1	2
	SM	1	3	2
	JL	2	1	3
	JW	2	3	1
	VL	1	2	3
	KH	2	3	1

Appendix Two : Percentage of words within each transcript classified as chunks by NSs and by each student

Student M

% of total text	underground (first)	food (second)	washing (third)
X: chunks for NSs	9	17	9
Y: ill-formed chunks within X	1	1	1
Z: Chunks for M	10	11	18
Overlap between X and Z	2	3	5

Student C

% of total text	bike and bars (first)	housemates (second)	movies (third)
X: chunks for NSs	15	29	12
Y: ill-formed chunks within X	1	0	1
Z: Chunks for X	5	4	7
Overlap between X and Z	0	3	4

Appendix Three : Extracts from M's transcripts, chunked by NSs and NNSs

Key: *italics* - words classified by the NNS as a chunk.

<u>underline</u> – words 2 or 3 out of the 3 NSs classified as a chunk

<u>underline</u> – words 2 or 3 NSs classified as an ill-formed chunk

NB: For reasons of space, only parts of one student's transcripts are reproduced here.

M's first talk in October – "underground"

1   ahh…first err I, I learned, learnt? (mmhmm I learnt) err (2.0) I should err.. I

2   <u>should be more positive</u>? (right) positive… in UK because ahh…when, when I

3   went to London err… last Sunday (mhmm) ahh (2.0) some, some of the

4   underground line (mm) line was *no service* (oh dear) ((speaker laughs)) *I was*

5   *<u>really surprised</u>* and, because it can, cannot be (mm) in Japan (mm) *you know*,

6   sun- in, in Sunday, on? (mm) <u>on Sunday</u> many, many people (mm) come to

7   London (mm) and go around some place (mm)... so everyone need to, need a

8   train (mm) so, but maybe four or five lines… was not, *no service* (mm) so…

9   I… I have to think err *<u>what I should do</u>* ((speaker laughs)) and no, I've never, <u>I</u>

10   <u>have never been</u> to London that, so, this was the first time I've been to London

11   (mm) so…

M's second talk in December – "food"

1   err… actually I (2.0) I have… <u>no problem</u> about living here in recent (mm) in

2   recently (mm) because *I <u>got used</u> to live* here (mm) and, but ahhh… <u>one</u>

3   <u>thing</u>… I just remember (mm) is err about food (mm)… ahhh *you know*

4   Western food and Japanese food are *<u>quite different</u>* (mm) and… so sometime,

5   some food is not my taste? (mm) and… especially err… the food… I have, I

6   have <u>at my college</u> (ahh!) ((both laugh)) is *<u>not so good</u>* ahh… it is very good,

7   I, I can get, I, I can have lunch <u>very cheaply</u>? (mm) but *the taste*

8   ((speaker laughs))

9   *is <u>not so good</u>* but I found… ahh… British people usually ahh…use salt?

10   (mmm) and pepper? (mm) before they eat (mm) and so… I think ahhh…

11   British people ahh… always try to **change** the taste (mm) ahh which (2.0)

12   are, is suit… (mm) for, for them? (mmm) so *I found the solution*

M's third talk in February – "washing"

9   well ahh… *you know*, I… I live with some… boys now… that's err…

10   sometimes, that's err… *caused some problem*… for me (speaker laughs) last

11   week ah… I used ah… dryer (mhmm) and…but I… I left my clothes inside,

12   in, in it (right) for a while, I forgot (mm)… and when I, so when I noticed

13   (mhmm) about it and I (2.0) I took, I…I tried to took, take? (mhmm) take

14   them and I found ahh… my clothes ahh… hold, hold, holded? … holded,

15   holded up with (held up? …ahh folded) folded (folded up) yeah, *folded up* on

16   it ((both laugh)) *I was so surprised* and… and ahh… one of my… ahh…

17   housemates told me ahh… Tim… ahh the, one of the, also one of the, my

18   flatmates did it so *I was so embarrassed* ((both laugh)) *I was so embarrassed*

19   and… but it's err… Japanese men ah… wouldn't, wouldn't do that… it's

20   because… very … ah… **private** thing (yeah) some underwear (2.0) mm they,

21   they know ahh… we Japanese women don't wanna… see ah… see women, do

22   them, so… ((both laugh)) that's our culture difference…

Appendix Four : Chunking the transcripts using WordSmith

Key: underline – words I classified as a chunk, guided by WordSmith Tools

                underline – words I classified as an ill-formed chunk, guided by JustTheWord

M – underground

1   ahh…first err I, I learned, learnt? (mmhmm I learnt) err (2.0) I should err.. I

2   should be more positive? (right) positive… in UK because ahh…when, when I

3   went to London err… last Sunday (mhmm) ahh (2.0) some, some of the

4   underground line (mm) line was no service (oh dear) ((speaker laughs)) I was

5   really surprised and, because it can, cannot be (mm) in Japan (mm) you know,

6   sun- in, in Sunday, on? (mm) on Sunday many, many people (mm) come to

7   London (mm) and go around some place (mm)... so everyone need to, need a

8   train (mm) so, but maybe four or five lines… was not, no service (mm) so…

9   I… I have to think err what I should do ((speaker laughs)) and no, I've never, I

10   have never been to London that, so, this was the first time I've been to London

11   (mm) so…

M – food

1   err… actually I (2.0) I have… no problem about living here in recent (mm) in

2   recently (mm) because I got used to live here (mm) and, but ahhh… one

3   thing… I just remember (mm) is err about food (mm)… ahhh you know

4   Western food and Japanese food are quite different (mm) and… so sometime,

5   some food is not my taste? (mm) and… especially err… the food… I have, I

6   have <u>at my college</u> (ahh!) ((both laugh)) is <u>not so good</u> ahh… <u>it is very good</u>,

7   I, I can get, I, I can have lunch <u>very cheaply</u>? (mm) but <u>the taste</u>

8   ((speaker laughs))

9   <u>is not so good</u> but <u>I found</u>… ahh… <u>British people</u> usually ahh…use salt?

10   (mmm) and pepper? (mm) before they eat (mm) and so… <u>I think</u> ahhh…

11   <u>British people</u> ahhh… always <u>try to **change**</u> <u>the taste</u> (mm) ahh which (2.0)

12   are, is suit… (mm) for, for them? (mmm) so <u>I found the solution</u>

M – washing

9   well ahh… <u>you know</u>, I… I live with some… boys now… that's err…

10   sometimes, that's err… <u>caused some problem</u>… <u>for me</u> (speaker laughs) <u>last</u>

11   <u>week</u> ah… I used ah… dryer (mhmm) and…but I… I left my clothes inside,

12   in, in it (right) <u>for a while</u>, I forgot (mm)… and when I, <u>so when I noticed</u>

13   (mhmm) <u>about it</u> and I (2.0) I took, I…<u>I tried to</u> took, take? (mhmm) take

14   them and I found ahh… my clothes ahh… hold, hold, holded? … holded,

15   holded up with (held up? …ahh folded) folded (folded up) yeah, <u>folded up</u> on

16   it ((both laugh)) <u>I was **so**</u> <u>surprised</u> and… and ahh… <u>one of my</u>… ahh…

17   housemates told me ahh… Tim… ahh the, <u>one of the</u>, also <u>one of the</u>, <u>my</u>

18   <u>flatmates</u> did it <u>so I was **so**</u> embarrassed ((both laugh)) <u>I was **so**</u> embarrassed

19   and… but it's err… Japanese men ah… wouldn't, wouldn't do that… it's

20   because… very … ah… **private** <u>thing</u> (yeah) some underwear (2.0) mm they,

21   they know ahh… <u>we Japanese women</u> don't wanna… see ah… see women, do

22   them, so… ((both laugh)) <u>that's our culture difference</u>…

Appendix Five : Fourth recording and M's comments

Key: underline – words I classified as a chunk, guided by WordSmith Tools

underline – words I classified as an ill-formed chunk, guided by JustTheWord

M's fourth talk in April – "room and spiders"

12   I… I hate ahh… all insects? (uhhuh) ((speaker laughs)) and spider, especially

13   spiders, especially British spiders ((speaker laughs)) ahh… I think, last, last

14   Sunday? (mm) I found a spider in my room (mm) and (2.0) it, which, which

15   has, has long legs (mm) and… is much bigger than Japanese one

16   ((speaker laughs)) so I was really surprised and… I didn't know… what to do

17   and how, how I kill it (mm) ((speaker laughs)) so ahh… I… asked to… ah one

18   of my flatmate to… kill it (mhmm)… and…er…and he said yes so he came in

19   my room… but it's, had disappeared (mm), we can't, we couldn't see it

20   anywhere so… ok thank you I said ((both laugh)) and… err ahh… I, and I

21   tried to start my work (mm) but I couldn't concentrate ((speaker laughs))

22   because I… ah… I knew… that a spider… is… in my room (mm) was in my

23   room anywhere, somewhere so… ahh… and so I

M's comments

"*always when I speak English I reassemble an English sentence like a puzzle*"

"*construction is really different*" *(of English and Japanese)*

"*when I think of a Japanese sentence and I try to speak in English I…I have to say… ahh… last word in Japanese*"

"*When I speak with native speakers… like now… I feel I have to speak quickly because other people are waiting*".

"*I need time to make a sentence*"