Tone and voicing perception in Walungge

Cathryn Bartram

Proceedings of Conference on
Language Documentation & Linguistic Theory 3

Edited by Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan

19-20 November 2011 School of Oriental and African Studies, University of London

Hans Rausing Endangered Languages Project
Department of Linguistics
School of Oriental and African Studies
Thornhaugh Street, Russell Square
London WC1H 0XG
United Kingdom

Department of Linguistics:
Tel: +44-20-7898-4640
Fax: +44-20-7898-4679
linguistics@soas.ac.uk
http://www.soas.ac.uk/academics/departments/linguistics

Hans Rausing Endangered Languages Project:
Tel: +44-20-7898-4640
Fax. +44-20-7898-4349
elap@soas.ac.uk
http://www.hrelp.org

ISBN: 978-0-7286-0398-1

This publication can be cited as:

Cathryn Bartram. 2011. Tone and voicing perception in Walungge. In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, 39-48. London: SOAS.

or:

Cathryn Bartram. 2011. Tone and voicing perception in Walungge. In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*. London: SOAS. www.hrelp.org/eprints/ldlt3_05.pdf

Tone and voicing perception in Walungge

CATHRYN BARTRAM
School of Oriental and African Studies

1. BACKGROUND INFORMATION

Walungge is an endangered Tibetan language spoken in the far northeast of Nepal. The Walungge speaking area is primarily the upper Tamur River and the Gunsa River in the north east of the Taplejung district of Nepal, and includes the villages of Olangchung Gola (the local name is Halung), Yangma, Gunsa and Lungthung. In this area there is estimated to be just over 1000 mother tongue Walungge speakers.

All Walungge language data and analysis is taken from Bartram (2011).

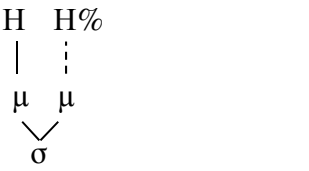
2. INTRODUCTION TO TONE IN WALUNGGE

The following is a brief introduction to Walungge tone. As is the case with many Tibetan languages, Walungge is tonal, with two lexical tones: high (H) and low (L). A word takes one underlying tone only. The following words illustrate the contrast between H (marked with an acute accent on the first syllable of the word) and L (marked with a grave accent):

(1)	*sáŋ*	'pine'	*sàŋ*	'saucepan'
	ɲíma	'cereal head'	*ɲìma*	'sun'
	ŋá	'drum'	*ŋà*	'1.SG'

The TBU (tone bearing unit) is the mora; the lexical tone attaches to the first TBU. In addition to this, Walungge has a non-lexical tone H%, which attaches to the final TBU of a word. In the particular case of a monomoraic word with H lexical tone, the non-lexical tone is L%. The combination of a lexical plus a non-lexical tone causes the realised pitch across the word to be either level, rising or falling. The following examples illustrate this, with pitch transcribed with numbers from 1 to 5 (5 being the highest):

(2)



sáŋ [saŋ ^{44}] 'pine'

sàŋ [saŋ ^{14}] 'saucepan'

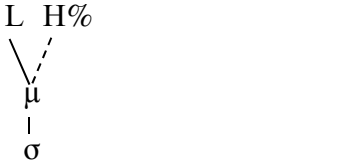
(3)

ɲíma [ɲima $^{44\ 44}$] ‘cereal head’

ɲìma [ɲima $^{11\ 44}$] ‘sun’

(4)

ŋá [ŋa ^{52}] ‘drum’



ŋà [ŋa ^{24}] ‘1.SG’

3. PERCEPTION OF TONE, VOICING, AND PHONATION

3.1. Introduction

Whilst voicing in Walungge is not phonologically contrastive, word initial plosives (and affricates) become phonetically voiced before low tone. The following pair of words illustrates this:

(5)

tá	[ta ^{52}]	‘horse’
tà	[da ^{24}]	‘bow’

In addition to this, phonation across a vowel closely follows the pitch melody, becoming breathier as the pitch falls and less breathy as the pitch rises.

This means that when a Walungge speaker hears a pair of words such as in (5) above, there are a number of acoustic cues available to enable the two words to be perceived as distinct, namely, whether the starting pitch is high or low, whether the overall pitch melody rises or falls, whether the initial consonant is voiced or voiceless, and how phonation changes through the vowel.

The sections that follow describe perception experiments that were carried out with mother tongue speakers of Walungge. The experiments investigate the relative salience between starting pitch, overall contour of the pitch, voicing, and phonation.

3.2. Experimental design

A series of binary choice word perception tests was devised. Subjects were played recordings and asked to identify the word they heard by choosing one of two pictures. Tonal minimal pairs were chosen and recordings were made of a mother tongue female speaker saying the words. Using these recordings, F0 and VOT were artificially manipulated to create sets of stimuli with different combinations of F0, VOT and phonation (see below for details). All acoustic manipulations were carried out using the computer program Praat v. 5.1.07. A number of different manipulated

values of F0 and VOT in the set of stimuli were chosen in order to have enough combinations for a thorough investigation, balanced against the need to prevent producing a set of stimuli so large that the experiment would take an unreasonably long time for subjects to complete. Each set of stimuli was presented to subjects in a randomised order, and all stimuli were presented twice in the experiment.

Because pitch melodies [52] and [24] are arguably the pair of pitch melodies with the greatest difference, it was decided to look first at pairs of bimoraic words, where the pitch melody of a word with underlying tone H is level rather than falling. Pairs of words starting with fricatives or sonorants were chosen, so that the cues of starting pitch, overall pitch melody and phonation could be considered before additionally considering initial consonant voicing. Pairs of words starting with plosives were then added to the experiment. Finally words with [52] versus [24] pitch melodies were added to the experiment.

3.3. Subjects

The experiments were carried out in the Walungge speaking area, and all subjects participating in the experiments were mother tongue Walungge speakers. All were born and raised in the Walungge area, and although some had lived away from the area for education, work, etc., all were currently living in the area. A total of 40 subjects participated: 38 of these gave usable results. Ages of subjects ranged from 13 to 61 years old, with the majority being between 20 and 40 years old. Approximately half were male and half were female. The subjects' level of education ranged from those who had never been to school to one person who had a BA degree. Most of the subjects had travelled outside the Walungge speaking area. For some, the extent of their travel was to Taplejung district centre. Others regularly made trips to Kathmandu, India, or Tibet. All spoke Nepali to some degree, and some spoke Lhasa Tibetan.

4. PERCEPTION OF STARTING PITCH AND OVERALL CONTOUR

4.1.Stimuli

In distinguishing between words which only differ in their underlying tone, e.g. *sáŋ* [saŋ ^{44}] 'pine' and *sàŋ* [saŋ ^{14}] 'saucepan', are Walungge speakers responding to the difference in high versus low initial pitch, or are they responding to the difference between a level versus a rising pitch melody? In order to investigate this, a set of stimuli with three distinct values of F0 was prepared, with the values chosen on the basis of acoustic analysis. The set of stimuli contrasted six rising or level pitch melodies based on these three values of F0:

shape of pitch melody	¯	–	_	ˊ	ˊ	ˊ
pitch melody "name"	3-3	2-2	1-1	1-2	2-3	1-3
starting F0 (Hz)	226	205	185	185	205	185
ending F0 (Hz)	226	205	185	205	226	226

Starting with a recording of the word *sáŋ* 'pine', the pitch was manipulated to create 6 recordings, each with one of the above 6 pitch melodies. To investigate the effect of phonation on word identification, a recording of the word *sàŋ* 'saucepan' was also manipulated in the same way. All 12 recordings were then manipulated to have identical durations, set to their average duration, resulting in a pair of stimuli for each pitch melody, identical except that one had originated from *sáŋ* and the other from *sàŋ*, preserving the slightly different phonation type of each. If subjects were responding in any way to the breathiness, this would be evident in the results of the experiment.

Because the level and rising pitch melodies occur on both monosyllabic and disyllabic words, recordings of the words *ɲìma* 'sun' and *ɲíma* 'cereal head' were manipulated in a similar way. Whereas the monosyllabic words were placed in frames when presented to subjects, the disyllabic words were played to people in isolation. In isolation, because of the lack of a reference point for the pitch, it is possible that greater salience is attached to the contour of the pitch melody, that is, whether it rises or is level.

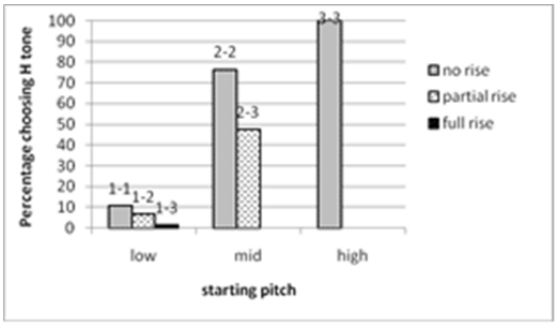
4.2. Results

The following charts give for each (manipulated) pitch melody the percentage of responses which identified stimuli carrying that pitch melody as the high tone word.

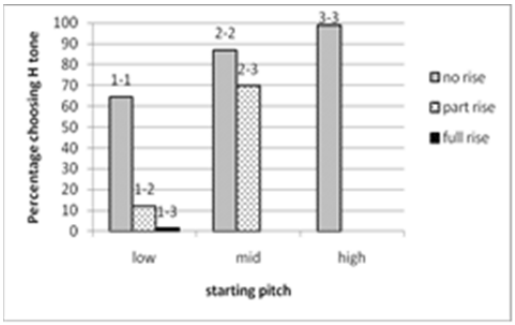
Figure 1

Percentage choosing H tone

stimuli in frames



stimuli in isolation



The high level pitch melody 3-3 and the low rising pitch melody 1-3 are the unaltered pitch melodies, and as expected virtually all subjects identified stimuli with pitch melody 3-3 as the H tone word *sáŋ* 'pine', and the stimuli with pitch melody 1-3 as the L tone word *sàŋ* 'saucepan'. This was the case regardless of whether the stimuli were in frames or in isolation.

When the pitch is low and level (1-1), subjects are presented with apparently conflicting acoustic cues. Whilst the low starting pitch is a cue that the item is an L tone word, a level pitch melody is normally associated with an H tone word. When the stimuli were presented in frames, the vast majority of subjects associated the low level melody with an L tone word. If people have to choose between the starting pitch and the shape of the melody, the starting pitch is by far the more salient cue, but

the shape of the melody still has a slight effect upon perception. However, when the stimuli were in isolation, the shape of the pitch melody became the more salient cue.

A mid starting pitch is a potentially ambiguous cue as to the tone of the word. In the case of stimuli presented in isolation with mid starting pitch, regardless of whether the pitch melody is level 2-2 or slightly rising 2-3, the majority of responses identified the stimuli as H tone words. This is different from framed stimuli, where the majority of responses for 2-2 were H tone, but not for 2-3.

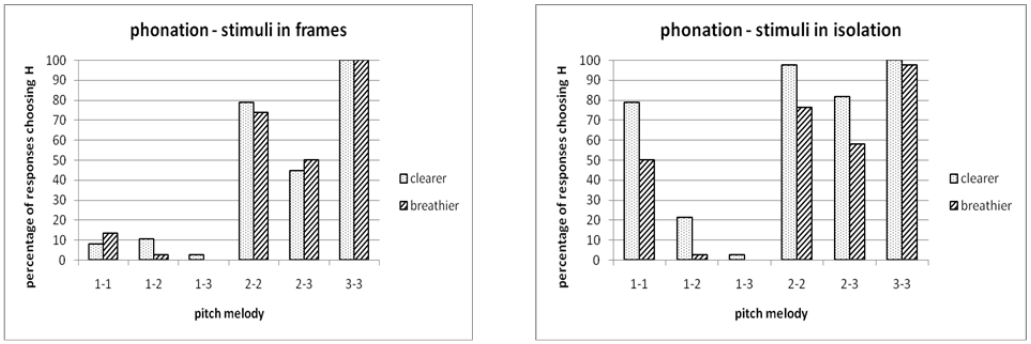
In summary, if there is a frame to provide a reference pitch (providing the starting pitch is not ambiguous) when presented with apparently contradictory cues the starting pitch is of greater salience than the shape of the melody. However when the starting pitch is ambiguous (mid), the contour of the melody becomes a salient cue in determining the tone of the word. When there is no frame giving a reference pitch, the pitch melody across a word is of greater salience than the starting pitch of the word. The starting pitch of the word still has an effect in that the higher the starting pitch the more likely it is for a listener to perceive that word as an H tone word. However, if it is a straight choice between the starting pitch and the overall pitch melody, then the shape of the pitch melody is more salient than the starting pitch.

5. EFFECT OF PHONATION

In the above experiments each pitch melody appeared twice. One occurrence was when the word that was manipulated to produce the pitch melody was originally an L word; the other occurrence was when the original word was an H word. Because the durations of the stimuli were manipulated to be identical, for each pitch melody the two stimuli were identical except for the phonation, which correlated with the original pitch melody of the word. Those stimuli which came from an L word had phonation which started breathier. Those stimuli which came from an H word had phonation which was less breathy.

The following graphs show the results for each pitch melody separated by breathiness:

Figure 2
Percentages for phonation



For stimuli in frames, phonation appears to play no role in word identification. There is no significant difference in the responses for those stimuli which had clearer phonation from those which had breathier phonation. However, for stimuli in isolation phonation does play a role in word identification. As seen in the above bar chart, for every pitch melody the percentage of responses identifying the stimuli as an H tone word was greater if the stimuli had clearer phonation than if they had breathier phonation. Furthermore, this difference was shown to be statistically significant. For words in isolation, the lack of a reference pitch from the surrounding words means that not only does the shape of the pitch melody across the word become the salient cue, but phonation also has a bearing upon how the word is perceived.

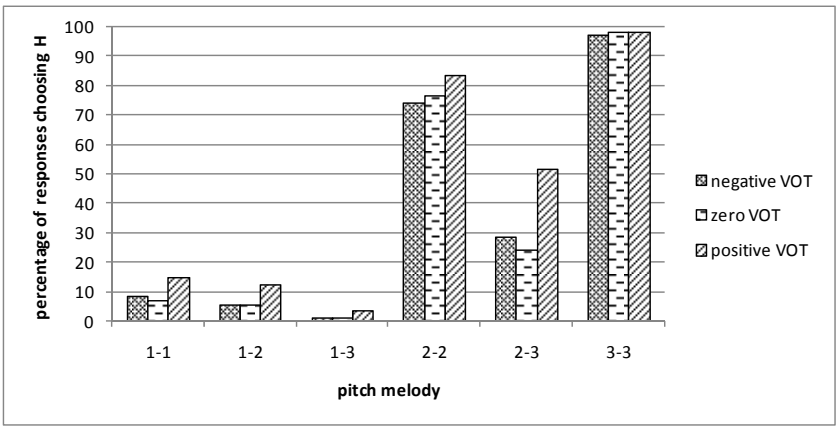
6. OBSTRUENT VOICING

Using the pair of words *cá:* 'wooden spoon' and *ɟà:* 'window', stimuli were created by manipulating the pitch to correspond to the same pitch melodies as used above. Durations were also manipulated in order for all stimuli to have identical durations. Following this, for each item the VOT was then manipulated to be one of 3 values: 20ms, zero, and -20ms, chosen according to acoustic analysis, giving 36 stimuli. The same manipulations were also carried out for the pair of words *túŋ* 'drink' and *dùŋ* 'hit', giving an additional 36 stimuli.

6.1. Results

The responses for all the stimuli in frames were collated and combined according to the (manipulated) pitch melody and the (manipulated) VOT. This gives 152 responses for each combination of VOT and pitch melody. The graph below gives a summary of the percentages of responses identifying the stimuli as H tone words.

Figure 3
Percentage of responses choosing H tone, for pitch and VOT



When the pitch melody is unambiguous (either 3-3 or 1-3) then VOT does not appear to play a role in word identification. For the high pitch melody 3-3, virtually 100% of responses equated this with an H tone word regardless of the VOT of the stimuli. Similarly, virtually 100% of responses equated the 1-3 pitch melody with an L tone word regardless of the VOT.

For pitch melodies other than 1-3 and 3-3, there are some differences in responses due to the VOT of the stimuli. In general the difference in responses correlates with the difference between positive VOT on the one hand and zero or negative VOT on the other hand. If the pitch starts low but does not do a complete rise (1-1 or 1-2), there is a very slight change in the responses as the VOT changes. But it has to be stressed that this is very slight, and the vast majority of responses matched a low starting pitch with L tone regardless of the VOT of the word.

In the case of a 2-2 pitch melody, the majority of responses identify this as H tone regardless of the VOT of the word. The 2-2 pitch is ambiguous because the level shape of the melody matches H tone but the height of the pitch is lower than it would normally be for H tone. This ambiguity results in there being a very slight VOT effect, with slightly more responses matching this to H tone when the VOT is positive.

The pitch melody where VOT does have a greater effect is 2-3. This is arguably the most ambiguous pitch melody: the starting pitch is neither high nor low, and the shape of the pitch melody neither rises to its full amount nor remains level. If the initial consonant had positive VOT approximately half the responses identified this as H tone. However, if the VOT was zero or negative, approximately three quarters of responses chose L.

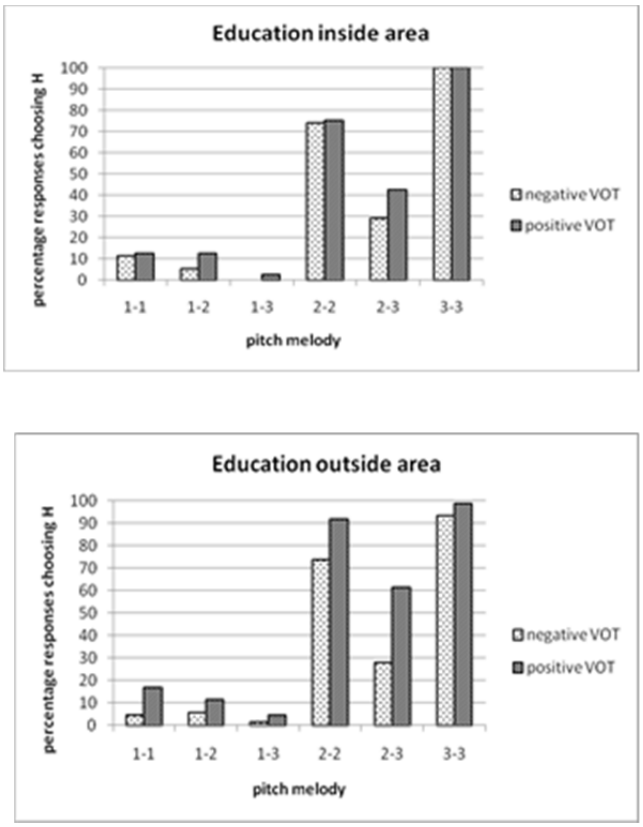
In summary, the salience of VOT is related to the ambiguity of the pitch melody. If the pitch melody is unambiguous, the VOT does not affect perception of words. Even if the pitch melody has some ambiguity associated with it, any VOT effect is generally only very slight. However, if the pitch melody is very ambiguous (as in the case of 2-3), then the VOT does become a salient cue in the perception of words.

6.2. Exposure to Nepali

One of the factors that is most likely to contribute to people's perception as the VOT changes is exposure to Nepali, or another language (such as English or some other Tibetan languages), which has a phonological voicing distinction. People are exposed to Nepali in a number of different ways, for example Nepali media, travel, education, proximity to Nepali speaking villages etc. But not all of these factors are easy to measure. One factor that is both easy to measure and is most likely to be having an effect is if people have been educated outside of the Walungge speaking area. The schools in the Walungge speaking villages are Nepali medium schools, but are only primary schools, and outside of the classroom Walungge is the language of village communication. For secondary school education and beyond, pupils have to board in a village or town outside of the language area; and all are sent to places where they have far greater contact with Nepali speakers (most commonly Lelep, Taplejung or Darjeeling). Approximately half the people taking part in the experiment had gone to school outside the Walungge speaking area.

In order to consider whether exposure to Nepali has an effect, the results of the above experiment were split according to whether or not people had gone to school outside the Walungge area.

Figure 5
Percentage choosing H; split by education



Pitch melodies 1-3 and 3-3 are the unambiguous pitch melodies with virtually everybody identifying 1-3 as L and 3-3 as H, regardless of VOT and regardless of whether people were educated inside or outside of the language area. From the initial VOT analysis above, the pitch melody where VOT made the greatest difference to the responses was 2-3. But notice from the above charts that the difference in responses according to VOT is greater for people who have been educated outside of the language area than inside the area. The analysis above also showed that VOT had an effect for pitch melody 2-2, but in the case of people educated inside the area VOT makes virtually no difference to the responses; it is only for people educated outside the area that VOT makes a difference.

For people who have only been educated inside the language area, VOT makes very little difference to the responses for any of the pitch melodies, even when the pitch melody is ambiguous (such as pitch melody 2-3). However, in the case of people who have been educated outside the language area VOT is making a difference to the responses to ambiguous pitch melodies.

6.3. Further results

Further sets of stimuli were created to extend the investigation of VOT perception. These included disyllabic words with a tone/VOT difference which were presented in isolation rather than frame, monosyllabic monomoraic words, and pairs of words which considered the interaction of phonation with VOT and pitch.

In summary, VOT made no difference for stimuli in isolation, either for people educated inside or outside the area. When words are in frames the starting pitch is of greater salience than the overall shape of the pitch melody. In the natural (unmanipulated) form of the language, VOT correlates with the tone, and hence the starting pitch. Because of the greater salience of the starting pitch as a cue to the meaning of the word, when the starting pitch becomes ambiguous, and in addition the shape of the pitch melody is ambiguous, VOT significantly affects the perception of the word, but generally only for people who have had sufficient exposure to Nepali. When words are in isolation the shape of the pitch melody across the word is of greater salience than the starting pitch. Because of the greater salience of the overall shape compared with the starting pitch, even when the pitch melody becomes ambiguous VOT has no significant effect upon the perception of the word. Phonation had no apparent effect upon perception.

7. DISCUSSION

One of the phonological issues in the analysis of Walungge is whether in a pair of words such as [ta ^{52}] 'horse' and [da ^{24}] 'bow' voicing is a phonological contrast and is determining the tone, or whether voicing is not a phonological contrast but is being determined by the tone. The conclusion presented in Bartram (2011) is that voicing is not phonologically contrastive for obstruents. However, the arguments are not completely clear-cut.

Now consider the diachronic process from a voicing contrast to a tone contrast. There is a point in this process for which Maran (1973) uses the term 'cognitivization' whereby 'the basic system of cognitive contrast is transferred from a segmental system of cues to a prosodic system. This event means, among other things, that it is now far simpler to conduct the cognitive process of perceiving and analysing linguistic data largely in terms of prosodic features of tone.' In going from a voicing contrast to a tonal contrast there is a change in perception of the prosodic cues and the segmental cues. The results of the perception experiments above suggest that this change has taken place, and give additional weight to the argument that voicing is not a phonological contrast in Walungge.

Exposure to Nepali (as measured by education outside the language area) affects perception, with a greater salience given to voicing in the instances where the pitch is ambiguous. The experiments described above for Walungge are similar to those described in Pearce (2007) for Kera. In a similar manner to Walungge, Kera has voicing of obstruents correlating with tone. Manipulations were carried out to create a series of stimuli with different combinations of VOT and F0. Pearce found that whilst those who had less exposure to French (e.g. village women) responded to F0

as the salient cue, those with greater exposure to French (e.g. town men) responded to the VOT as the most salient cue. Pearce concluded that whilst diachronically the tone contrast in Kera has arisen from a voicing contrast, in the synchronic form of the language (as spoken by those without exposure to French) voicing is not contrastive. However, she also concluded that exposure to French is causing voicing to be reanalysed as contrastive in the town dialect of the language. It is possible that in Walungge a similar reanalysis could happen. However, from the perception experiments it appears that Walungge is not yet at that point. Even for those exposed to Nepali, it is only in the cases where the pitch is ambiguous that VOT becomes salient. In all other instances, pitch is still the most salient cue.

REFERENCES

Bartram, Cathryn V. 2011. *An investigation of tone in Walungge.* Ph.D. dissertation, SOAS

Maran, L. R. 1973. On becoming a tone language: a Tibeto-Burman model of tonogenesis. In L. Hyman (ed.), *Consonant types and tone,* Southern California Occasional Papers in Linguistics, Linguistics Program USC. LA, no 1.

Pearce, Mary. 2007. *The interaction of tone with voicing and foot structure: Evidence from Kera phonetics and phonology*. Ph.D. dissertation, UCL.