Discourse functions of duration in Mandarin: resource design and implementation

Shu-Chuan Tseng*, **Dafydd Gibbon**†

*Academia Sinica, Taipei, Taiwan, tsengsc@gate.sinica.edu.tw
† Univerität Bielefeld, gibbon@uni-bielefeld.de

Abstract

A dedicated resource, consisting of annotated speech data, tools, and workflow design, was developed for the detailed investigation of discourse phenomena in Taiwan Mandarin. The discourse phenomena have functions which are associated with positions in utterances, and temporal properties, and include *discourse markers* ('NAGE', 'NA', e.g. 'hesitation', 'utterance initiation'), *discourse particles* ('A', e.g. 'utterance finality', 'utterance continuity', 'focus', etc.), and *fillers* ('UHN', 'hesitation'). The distribution of particlers in relation to their position in utterances, and the temporal properties of particles are investigated. The results of the investigation diverge considerably from claims in existing grammars of Mandarin with respect to utterance position, and show in general greater length than for regular syllables. These properties suggest the possibility of developing an automatic discourse item tagger.

1. Introduction

A dedicated resource of data and tools was developed for the detailed investigation of discourse phenomena in Taiwan Mandarin, consisting of annotated speech data, tools, and workflow design. The discourse phenomena which were of special interest include *discourse markers* ('NAGE', 'NA', e.g. 'hesitation', 'utterance initiation'), *discourse particles* ('A', e.g. 'utterance finality'), 'utterance continuity', 'focus', etc.), and *fillers* ('UHN', 'hesitation'). Study of these phenomena in the resource has revealed that the functions of the discourse phenomena are associated with utterance location and duration.

The specific research task is based on the initial observation that discourse phenomena are associated with specific temporal properties, which leads to the suggestion that duration patterns in the immediate contexts of these phenomena can (a) assist in speech recognition, (b) guide naturalistic speech synthesis. The paper deals with the corpus, Mandarin discourse particles and markers, the temporal analysis procedure, the results and finally a summary and conclusion.

2. Mandarin Conversational Dialogue Corpus (MCDC)

For the analysis of temporal properties of discourse phenomena, the following requirements were specified:

- an annotated corpus of recordings of spontaneous speech,

- suitable annotation mining tools,

- annotations with at least the granularity of syllables on a phonetic tier,

- discourse particle classes on a lingusitic tier.

The use of existing Taiwan Mandarin resources such as TWPTH, LDC 1998, (Duanmu et al., 1998) was considered. This corpus is available with transcriptions but is not provided with temporal annotations, although the corpus is suitable for investingating larger granularity items such as intonation units and repairs.

It proved to be necessary to design and create a dedicated annotated corpus for the discourse phenomenon duration task (MCDC: Mandarin Conversation Dialogue Corpus, (Tseng, 2004), (Tseng, 2005)), which was recorded from March to July 2001.

Currently the MCDC consists of 8 hours of annotated recordings of spontaneous conversational speech by pairs of speakers. The speakers were recruited from a random sampling of 1000 Taipei citizens who were invited to the recording sessions; 60 of those contacted responded to the request and were recruited. None of the speakers knew any of the others before the conversations, permitting the conversations to be classified as natural conversations between two strangers. The conversation partners were asked to introduce themselves at the start of the conversation; the rest of the conversation was left to the conversation partners. The total length of the 30 conversations is 25.6 hours, with an average length of 50 minutes.

The annotation system is designed to mark up detailed spontaneous speech phenomena:

1. disfluency,

2. sociolinguistic phenomena,

3. particular pronunciation.

A detailed description of the collection, processing and annotation of the MCDC is available.[1]

All sound files were segmented and stored in stereo files in order to build a multimodal database together with the metadata of the transcription texts. The total size of processed sound files is 18 GB. With the help of Translist, 8 conversations (136,000 characters) were completely transcribed and annotated. Sound files and transcripts of the eight conversations are available for public use. The corpus data have been licensed to the Association for Computational Linguistics and Chinese Language Processing.[2]

---

[1]At <http://mmc.sinica.edu.tw>
[2]At <http://www.aclclp.org.tw/use_mat.php#mcdc>

3. Mandarin discourse particles and markers

3.1. Classification

The background to discourse particles and markers in general and for Mandarin are discussed in (Schiffrin, 1987) and in (Chu, 2002; Liu, 2005), respectively.

In the transcription convention developed for the MCDC, we divide tokens of particle–like items in Taiwan Mandarin into four main groups according to the shared discourse functions, for example:

1. turn–keeping,

2. gap–bridging in language planning and production,

3. phatic response to maintain the communication channel.

The discourse items differ in a number of respects. In the writing system of Mandarin, some particles are conventionalized and provide a rather fixed interpretation in terms of both pragmatic and lexical meaning.

To take EN as an example, whenever an EN–like sound, e.g. UHN, is used for hesitation, it will be immediately re–interpreted as EN rather than a filler UHN by a native speaker, because EN has a principle meaning of hesitation, although EN and UHN have clearly different pronunciations and different pragmatic functions: the item UHN is often used as a filler, whereas EN is used for a prolonged (emphasized) hesitation.

The four groups of particle–like items are briefly introduced below. In the MCDC transcription, they are all transcribed in upper case characters (Tseng 2005) to enable easy differentiation from other categories of words.

3.2. Discourse particles with corresponding characters in the writing system

A number of discourse particles in Taiwan Mandarin are written in fixed characters (here in a transliterated form):
A, AI YA, AI YOU, BA, E/EP, EN, HAI, HE, HEI, HWA, LA, LIE/LEI, LO, MA, NOU/NO, O, OU, WA, WA SAI, YE, YI, YOU.
Most of them are monosyllabic, only a few them are disyllabic. Because particles conventionally do not have substantial lexical meaning, definitions in sound and form are not as strict as the other lexical entries. For instance, E has a pronunciation variant in spontaneous speech, i.e. EP, which has a bilabial coda after the vocalic part E. Also, NOU has the variant NO: in the pronunciation of NO, the vocalic part U is reduced. LIE–LEI, E–EP NOU–NO in the list are written with the same characters, whereas O can be written with more than one character.

3.3. Discourse particles without corresponding characters in the writing system

Not all discourse particles in Taiwan Mandarin have conventional fixed characters. For instance, frequent particles such as AI YE, EI, HEN, HON, NEI, and ON are normally written in traditional Zhuyin transcription in texts.

3.4. Discourse particles stemming from Taiwanese Southern–Min

Particles such as EIN, HAN, HEIN, and HO originate from the main dialect spoken in Taiwan, the Southern–Min dialect. Another Chinese dialect spoken in Taiwan, Hakka, is much less dominant than Southern–Min, and discourse particles stemming from Hakka are rarely adopted in the use of Mandarin.

3.5. Fillers

Table 1: Filler types.

Fillers–1	Fillers–2	Fillers–3	Fillers–4
UHN	UHM	NHN	MHM
UHNN	UHMM	NHNN	MHMM
UHNHN	UHMHM	NHNHN	MHMHM

The main function of fillers in Mandarin is to respond to the conversation partners. To take Fillers–1 in Table 1 as an example, the first occurrence UHN is monosyllabic, the second UHNN has a prolonged nasal part and the third UHNHN is disyllabic. The difference between UHN in Table 1 and EN in the list of particles with corresponding characters in the writing system is that UHN is used to respond to the conversation partner and EN is used mainly for hesitation. Furthermore, fillers in Taiwan Mandarin are not represented in the writing system for Mandarin, i.e. they are not written in fixed characters.

3.6. Markers

Table 2: Discourse Markers

Transcription	Original Meaning
NA	that (determiner)
NE	that (determiner)
NAGE	that + classifier
NEGE	that + classifier
NEIGE	that +classifier
SHENME	what
ZHEGE	this +classifier

Discourse markers are defined in our data as *lexical items which are independent lexical entries*. When used in spoken conversation, their function is pragmatic only. Table 2 lists the most frequent discourse markers found in Taiwan Mandarin. Interestingly, there is a clear difference between NA and NAGE: NA in Table 2 is related to the whole sentence, and occurs frequently in utterance–initial position; on the other hand, NAGE is mainly related to lexical or phrasal processing, and occurs mostly in utterance–medial position. The principles of disambiguation for ambiguous tokens, e.g. as discourse markers or determiners, are mainly based on the word following the ambiguous item. When the subsequent word is a proper noun, then it is more likely to be a discourse marker, because determiners do not precede proper nouns in Mandarin.

4. Temporal analysis procedure

For the specific task of analysing temporal properties of discourse phenomena, two tiers were initially defined:

1. syllables,

2. the presence/absence of discourse phenomena.

A discourse phenomenon lexicon was created, from which a third tier of discourse phenomenon categories is automatically inserted into the annotation, following the procedure described in (Gibbon et al., 2003).

A multi–stage workflow specification for the annotation mining procedure was developed for processing the corpus:

1. Production of annotation with forced alignment segmentation and labelling using the HTK package. (An extension of the original tabular HTK/esps format was used as input to computational processing; see below.)

2. Format conversion and manual evaluation of annotation with WaveSurfer.

3. For more general use, the annotations were reformatted in the standard TASX XML format (Gut and Milde, 2003) and checked for syntactic correctness, using an existing conversion toolkit.

4. Further format conversion into the Praat phonetic workbench format was performed, for phonetic processing with Praat.

5. A toolkit was prepared for mining the annotations and preprocessing duration sequences:

   (a) duration preprocessing tools for extracting duration sequences from the annotations,

   (b) tools implementing a number of operations and transformations for further analysis.

6. Annotation mining was performed with the toolkit.

The operations performed on the extracted duration sequences include:

1. z–score duration normalisation following (Hirst and Bouzon, 2005);

2. first derivative as normalised differences between neighbours, for identifying rhythmic and prominence patterns;

3. locally normalised first derivative, i.e. the Pairwise Variability Index, PVI (Low et al., 2000);

4. a further transformation which reduces the first derivative to binary values based on the polarity of the difference.

The output of the toolkit was collated for further processing in tabular form as an extension of the original tabular esps format.

In order to develop initial task–directed hypotheses, the output of the toolkit resource was 'eyeball–evaluated' by inspection of visualisations of the duration patterns and their transformations, in order to develop the initial hypotheses for computational annotation mining and statistical evaluation.

Two main properties of the discourse phenomena were examined:

1. position in the utterance,

2. duration properties.

5. Results

5.1. Discourse item in relation to utterance position

5.1.1. Fillers

Table 3: Durations of discourse items relative to position in utterance.

Type	Tot	U_init	U_fin	U_med
Discourse markers	*50	7	1	42
Discourse particles	136	24	27	85
Fillers	51	34	3	14

* NAGE, ZHEGE, NEIGE are disyllabic.

Fillers occur more often in utterance–initial position. Discourse particles are more often found in utterance–medial positions; utterance–final and utterance–initial in similar percentage. Discourse markers are used more often in utterance–medial positions; the second most likely position is utterance–initial.

In the literature on discourse items in Mandarin, particles are normally found in the utterance–final positions. However, in spontaneous Mandarin, the utterance may not be a suitable unit when discussing the position and location of discourse items.

5.1.2. Discourse markers

The most frequently used discourse markers are 'NAGE' and 'NA'. In the corpus, we also found 'ZHEGE' and 'NEIGE'. 'NA' appears mostly in utterance–initial or clause–initial positions, and 'NAGE', 'ZHEGE', and 'NEIGE' more likely in utterance–medial positions. This may have to do with the original meaning of 'NAGE', in which 'GE' is a classifier occurring frequently within sentences, especially before nouns. This tendency is clearly supported by our data.

5.1.3. Utterance–initial and utterance–final particles

Utterance–initial discourse particles are 'ON', 'E', 'O', 'EN', 'HEN', 'EI' and 'A'. Utterance–final discourse particles are 'MA', 'LA', 'BA', 'A', 'HON' and 'E'. In general, we can say that 'MA', 'LA', and 'BA' are solely utterance–final particles. The positionally most variable particles are 'E' and 'A', which can be used in utterance–final, utterance–initial and utterance–medial positions, depending on their pragmatic functions (Liu, 2005).

5.2. Temporal properties of discourse items

In general, the discourse items which have been investigated are longer than the normal syllable duration. Fillers are especially long in all utterance–positions when used for

hesitation, but for other categories of discourse items, this is not the case.

Utterance–initial discourse markers are, interestingly, very short. The 'NA' tokens in the beginning of utterances indicate utterance or turn initiation rather than hesitation. However, the utterance–medial discourse markers have longer than average syllable duration, and may also have the function of hesitation.

The analysis shows that utterance position, duration and pragmatic function may relate closely to each other. utterance–final and utterance–medial discourse particles have only slightly longer than average syllable duration, but utterance–initial discourse particles are much longer, and may function as an emphasized utterance–initiation.

The variation of duration seems to be reflected better by the z–scores, suggesting that if a 'syllable–window' can be developed for checking the changes of z–score during the utterance it may be possible to calculate weights for identifying discourse items.

6. Summary and conclusion

The present contribution describes the MCDC corpus, and the workflow and toolkit specifications for annotation mining, and reports on the use of the duration processing tools. The annotated corpus and the toolkit wil be released for general use after further development and evaluation.[3]

The results show that fillers tend to occur in utterance–initial position and discourse particles and markers more in utterance–medial position. Discourse markers are seldom used in utterance–final positions. In general, 'NA' is more likely to occur in utterance–initial or clause–initial position, and 'NAGE', 'ZHEGE', and 'NEIGE' in utterance–medial position. This may have to do with the original meaning of 'NAGE', in which 'GE' is a sentence–internal classifier which occurs especially before nouns, a tendency which is supported by the data.

It is often claimed that discourse particles in Mandarin are utterance–final, but in the corpus they occur in different utterance positions. Frequent utterance–final discourse particles are 'MA', 'LA', 'BA', 'A', 'HON', and 'E', and frequent utterance–initial discourse particles are 'ON', 'E', 'O', 'EN', 'HEN', 'EI', and 'A': 'MA', 'LA' and 'BA' are utterance–final particles, while 'E' and 'A' are variable and occur in utterance–final, utterance–initial and utterance–medial positions.

In general, discourse item duration is longer than normal syllable duration:

1. Fillers are especially long in all utterance positions, and are mainly used to mark hesitation.

2. Utterance–initial discourse markers are, interestingly, very short; 'NA' at the beginning of utterances indicates a signal of utterance or turn initiation rather than a hesitation.

3. Utterance–medial discourse marker duration is also longer than average syllable duration, and may also indicate hesitation.

4. Utterance–final and utterance–medial discourse particles have only slightly longer than average syllable duration, but utterance–initial discourse particles are much longer, and may have the function of emphasised utterance–initiation.

In our statistical results, z-scores appear to reflect the variation of duration best. We will further examine the changes of z-scores along the utterance in order to find indices for the locations of the discourse phenomena, and investigate the use of our results in developing the duration–based component of a discourse phenomenon tagger.

7. References

Chauncey C. Chu. 2002. Relevance theory, discourse markers and the Mandarin utterance–final particle A/Ya. *Journal of the Chinese Language Teachers Association*, 37(1):1–41.

S. Duanmu, G. H. Wakefield, Y. P. Hsu, G. Cristina, and S. P. Qui. 1998. Taiwanese Putonghua Speech and Transcript Corpus. *Linguistic Data Consortium*.

Dafydd Gibbon, Benjamin Hell, Felix Sasaki, and Thorsten Trippel. 2003. Acquiring lexical information from multilevel temporal annotations. In *Proceedings of EUROSPEECH 2003*, Geneva.

Ulrike Gut and Jan-Torsten Milde. 2003. Annotation anad analysis of conversational gestures in the TASX environment. *Künstliche Intelligenz*, 17.

Daniel Hirst and Caroline Bouzon. 2005. The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English). In *Proceedings of INTERSPEECH–EUROSPEECH 2005, Lisbon*.

Vincent J. Liu. 2005. *A corpus–based study of particle A in spontaneous Mandarin*. Ph.D. thesis, Graduate Institute of Linguistics, Fu–Jen University.

Ee Ling Low, Esther Grabe, and Francis Nolan. 2000. Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4):377–401.

Deborah Schiffrin. 1987. *Discourse Markers*. Cambridge University Press, Cambridge.

Shu-Chuan Tseng. 2004. Processing spoken Mandarin corpora. *Traitement automatique des langues, Special Issue on Spoken Language Processing*, 45(2):89–108.

Shu-Chuan Tseng. 2005. Syllable contractions in a Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics*, 10(1):63–83.

---

[3]At <http://mmc.sinica.edu.tw>