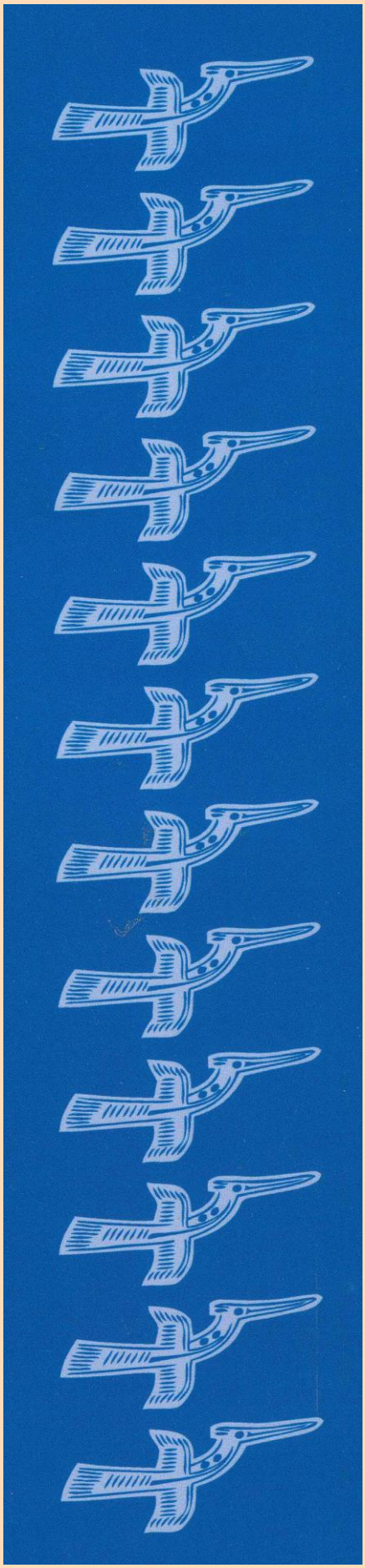
Mon-Khmer Studies

VOLUME 43.2



The journal of Austroasiatic languages and cultures

1964—2014 50 years of MKS

Author: *NGUYỄN Anh-Thư*
Title: ***Acoustic correlates of rhythmic structure of Vietnamese narrative speech.***
Pages: 1-7

Volume 43.2 Editors:
Paul Sidwell
Brian Migliazza

ISSN: 0147-5207
Website: http://mksjournal.org

Published by:

Mahidol University (Thailand)

SIL International (USA)

Acoustic correlates of rhythmic structure of Vietnamese narrative speech

Anh-Thư T. Nguyễn
Mountain Creek, Sunshine coast, Australia

Abstract
This paper reports a study on the acoustic realization of the rhythmic structure of Vietnamese narrative prose speech. Eight speakers of Saigon dialect read part (one page) of a short story. Acoustic measurements including duration and intensity were taken for every syllable of the excerpt. The syllables were labelled into four types: s is for a monosyllable that stands alone by itself; s0 is for the first syllable of a three-word phrase/chunk in which it is the modifier of a bisyllabic word (e.g., *cái* ý kiến, *thật* âu yếm); s1 and s2 are the first and second syllables of a bisyllabic word/chunk (e.g., ý kiến, âu yếm). The one-syllable, two-syllable and three-syllable units/chunks were also labelled utterance final and utterance non-final. The results showed that for both utterance medial and utterance final chunks, the monosyllable significantly had longer duration and stronger intensity than the other syllable types. Within bisyllabic words, the second syllables had longer duration and stronger intensity than the first syllables. Within three-syllable phrases/chunks, the first syllable of a three-word phrase/chunk was not significantly different from the first syllable of the bisyllabic word and the second syllables of the bisyllabic words/chunks also had longer duration and stronger intensity than the preceding first syllables. This result suggests an iambic pattern of acoustic prominence of bisyllabic and trisyllabic words/phrases in narrative speech.
Key words: Acoustic correlates, rhythmic structure, narrative speech.
ISO 639-3 language codes: vie

1. Introduction

Vietnamese is a contour tone language and has no system of culminative word stress; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson, 1965; Nguyễn Đăng Liêm, 1970). Duration, intensity, full tonal realisation of accented syllables have been observed to be important parameters for describing stress in Vietnamese (Đỗ, 1986; Chaudhary, 1983; Hoàng & Hoàng, 1975; Gsell, 1980). Regarding the stress patterning in utterances, it is generally agreed by some researchers that there is an alternating pattern of strong and weak syllables. Thompson (1965) stated that the majority of the syllables have medium stress. In a sequence of syllables, alternating ones are slightly louder (but not in a distinctive manner): "each pause group has at least one heavy stress and weak stresses are fairly frequent in rapid passages, rarer in carefully speech" (p. 50). Jones and Huỳnh (1960) stated that "normally the stresses in a Vietnamese utterance are conditioned by the junctures," and regarded the fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress. Consistent with Jones and Huỳnh's observation, it is remarked by Cao (2003) that due to the demarcative function of stress/accent in Vietnamese, native listeners tended to hear a juncture after a stressed syllable even though there is no such pause in reality as examined by spectrograms. In a recent study, Schiering, Bickel and Hildebrandt (2010) remarked that "Vietnamese provides ample evidence for a genuine stress domain that is preferably disyllabic and maximally trisyllable. Within this domain, stress is realised on the final syllable in the default case. Crucially, this domain is computed irrespective of the morphosyntactic status of its constituent syllables, i.e. stress phonology does not distinguish between a word-level and a phrasal-level of prosodic structure. Metrically, polysyllabic words are thus indistinguishable from other combinations of syllables. Since the most complex structures which are referenced by the rules for iambic rhythm are phrasal, stress may most adequately be attributed to the prosodic domain of the Phonological Phrase."(p.673).

In recent studies on more carefully phonetically controlled and specialized sets of Vietnamese disyllabic compounds and reduplications, Nguyen and Ingram (2007a, b) have found that there was at least a phonetic tendency for the right hand element of a disyllabic compound

word to be more prosodically prominent by a number of relevant phonetic measures: greater tonal f0 range, higher intensity, greater duration of the second syllable, and formant measurements indicative of more centralized vowel nuclei (vowel reduction) on the first syllable. Nguyen (2010) investigated the rhythmic patterns in Vietnamese polysyllabic words by examining the rhythmic patterns and their acoustic correlates in polysyllabic reduplicative words (2-,3-,4-,5-,6- syllable pseudo-words). The results showed that there is a tendency of syllable coupling indicated mainly by syllable duration pattern and supported by the native listeners' perception results, suggesting that polysyllabic words in Vietnamese tend to be parsed into bi-syllabic iambic feet with a rightward or retrograde rhythmic pattern. In a recent study, Nguyen (2013) examined the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poems made up of three-word, five-word, six-word, seven-word, and eight-word lines. The acoustic analysis showed that the duration and intensity results mirror each other in indicating a strong iambic pattern of prominence, supporting the literature that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn, 1985). Nevertheless, the perception results showed that listeners'parsing patterns, though to some extent reflect the acoustic patterns, do not strongly correlate with the acoustic results. This study is a follow-up of the results found in Nguyen (2010) and Nguyen (2013) that polysyllabic units of speech in Vietnamese tend to be parsed into bi-syllabic iambic pattern as indicated by the examination of duration and intensity patterns of syllables in narrative prose speech.

2. Experiment

2.1. Linguistic materials

In other to pursue the aim of the study, part (one page length) of a short story titled *Tôi đi học* (I went to school) by Thanh Tịnh was used. The excerpts consisted of 372 syllables in total. The excerpt is in the appendix.

2.2. Subjects

Eight speakers of the Sài Gòn dialect (4 males, 4 females) who came from Hồ Chí Minh city participated in the study. They were either visitors or newly arrived immigrants to Australia and had been in Australia from 2 weeks to 4 years. Their age ranged from 38 to 70 years. Their education levels ranged from high school to higher degrees.

2.3. Procedure

Subjects were given the short story in print to practice reading before the recording. They were asked to read the excerpts in a natural narrative manner. Recordings were made in a quiet room using sound recording and editing computer software PRAAT (Boersma and Weenink, 2007) at 22050 Hz sampling rate.

2.4. Measurement

The acoustic parameters measured included syllable duration (ms) and syllable intensity (dB). Temporal variations of F0 and tonal shapes are obvious components but will not be treated here due to the nature of the linguistic material. That is, the examination of tones requires a comparison and contrast of constant segmental compositions of the target linguistic materials which cannot be met by the nature of the short story. Peak intensity (dB) in syllables and syllable duration (ms) were measured manually via Praat (Boersma and Weenink, 2007).

The one-syllable, two-syllable and three-syllable units/chunks of speech were segmented by the researcher on the basic of auditory signal and spectrogram: cues for segmentation included final lengthening, pause, F0 reset and auditory rhythmic perceptual cues. An example of the segmented chunks was presented in Appendix 2. The syllables were labelled in a Praat Textgrid into four types by the researcher based on auditory signal and spectrogram: s is for a monosyllable that stands alone by itself (e.g. *và* [and], *lại* [again], ấy[that], tôi [I]); s0 is for the first syllable of a three-word phrase/chunk in which it is the modifier of a bisyllabic word (e.g., *cái* ý kiến, *thật* âu yếm); s1 and s2 are the first and second syllables of a bisyllabic word/chunk (e.g., ý kiến, âu yếm). In addition, utterances were also identified based on the punctuations in the text (periods and commas) together

with the silent pauses and pitch reset in the speech signal. The one-syllable, two-syllable and three-syllable units/chunks were also labelled utterance final and utterance non-final. The utterance has been proposed as the largest unit in the prosodic hierarchy: It is the largest span of application of phonological rules (Selkirk, 1978, 1980; Nespor & Vogel, 1986; Hayes, 1989) and its boundaries are sometimes said to be the location of non-hesitation pauses (Hayes, 1989). This unit often corresponds to a single syntactic sentence, but can include two or more sentences joined into a single higher-level sentence (Selkirk, 1978). It is noted that all of the utterances in the narrative were found to be followed by silent pauses and/or with pitch resets.

2.5. Analysis

There were in total 2616 syllables (327 syllables per excerpt x 8 speakers). A mixed (fixed and random) effects analysis of variance (ANOVA) model, using the restricted maximum likelihood method (REML) to estimate variance components was used to statistically analyse the data. The dependent variables were syllable duration and intensity. The fixed effects were syllable positions (s, s0, s1, s2), segmented chunks (one-syllable, two-syllable, and three-syllable chunks) and chunk positions (utterance final and utterance non final). The random effect was speakers and items. Tukey post-hoc tests were carried out to determine the significant differences among levels of the main effects when necessary.

The use of REML overcomes the potentially serious deficiency of the ANOVA-based methods which assumed that data are sampled from a random population and normally distributed. REML also avoids bias arising from maximum likelihood estimators in which all fixed effects are known without errors, consequently tend to downwardly bias estimates of variance components. Moreover, REML can handle unbalanced data. The data analysis was carried out using SAS program.

2.6. Results

2.6.1. Prosodic units

Majority of the narrative speech was segmented as two-syllable units (60.4%), while three-syllable units accounted for 23.4% and one-syllable unit only 16.2%.

2.6.2. Duration

The three-way mixed effect ANOVA (syllable positions x segmented chunks x chunk positions) on syllable duration showed a significant effect for the main factor syllable positions: $F(5, 2931) = 9.12$, $p < .0001$, chunk positions $F(1, 2931) = 69.77$, $p<.0001$, but no significant effects for segmented chunks $F(2, 2931)= 0.03$, $p= 0.5$ ns. The interactions syllable positions x segmented chunks x chunk positions: $F(1, 2931) = 0.9$, $p =.09$ ns and other interactions were not significant.

A post-hoc Tukey test on the significant differences among levels of the main factor of syllable positions (figure 1 below) showed that for both utterance medial and utterance final positions, the monosyllable significantly had longer duration than the other syllable types. Within bisyllabic words/ chunks, the second syllables had longer duration than the first syllables. Within three-syllable phrases/chunks in non-final position the first syllable (s0) of a three-word phrase was not significantly different from the first syllable (s1) and second (s2) of the bisyllabic words and within the bisyllabic words the second syllable (s2) was significantly longer than the first syllable (s1). In utterance final position, the first syllable (s0) of a three-word phrase was significantly different from the first syllable (s1) of the bisyllabic words and within the bisyllabic words the second syllable (s2) was significantly longer than the first syllable (s1). Generally, the results are very robust for both utterance final and non-final positions. Utterance final syllables are also shown to be marginally longer than those at utterance non-final position, indicating a final lengthening effect.

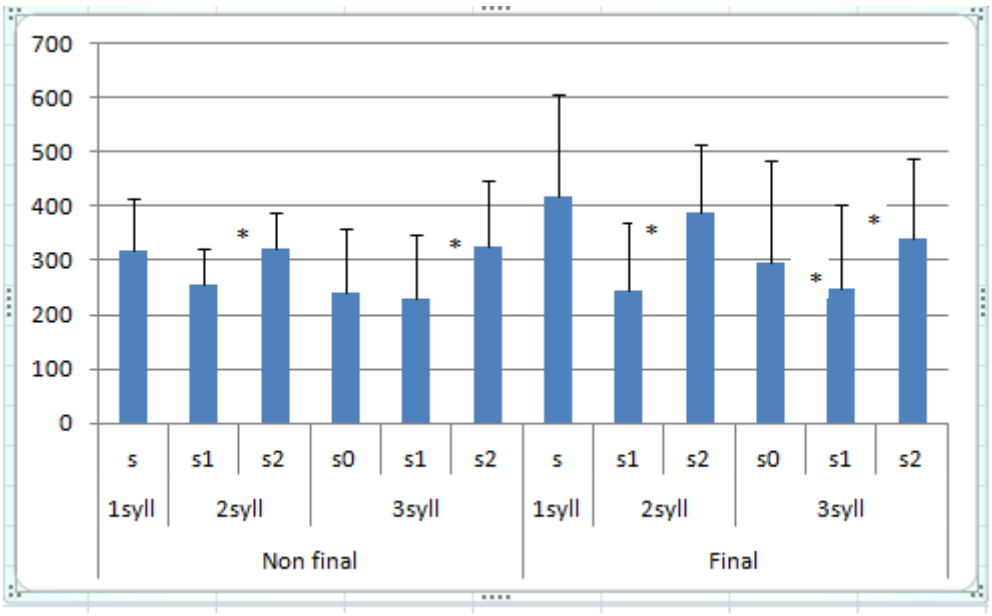


Figure 1: Mean duration (ms). s, s0, s1 and s2 are syllable positions in segmented chunks. 1syll, 2syll, and 3syll: number of syllables in chunks.
Non final and final are positions of chunks in utterances. The symbol * means p < 01

2.6.3. Intensity

The three-way mixed effect ANOVA (syllable positions x segmented chunks x chunk positions) on syllable intensity showed a significant effect for the main factors syllable positions: F (5, 2931) = 6.16, p < .0001, and chunk positions F (1, 2931) = 68.5, p = <.0001, but no significant effect for segmented chunks F (2, 2931) = 0.84, p=0.4 ns. The interaction syllable positions x segmented chunks x chunk positions: F (1, 2931) = 1.31, p =0.2 ns. and other interactions were not significant.

A post-hoc Tukey test on the significant differences among levels of the main factor of syllable positions (figure 2 below) showed that in both utterance non-final and final positions, the monosyllable significantly had stronger intensity than the two other syllable types (s0 and s1). Within bisyllabic words/chunks, the second syllables had stronger intensity than the first syllables. Within three-syllable phrases/chunks, the first syllable (s0) of a three-word phrase was not significantly different from the first syllable (s1) of the bisyllabic words and the second syllable (s2) also had stronger intensity than the first syllable (s1). The results also show that intensity value of the utterance–final syllables is lower than that of utterance non-final ones, indicating an intensity declination at utterance final.

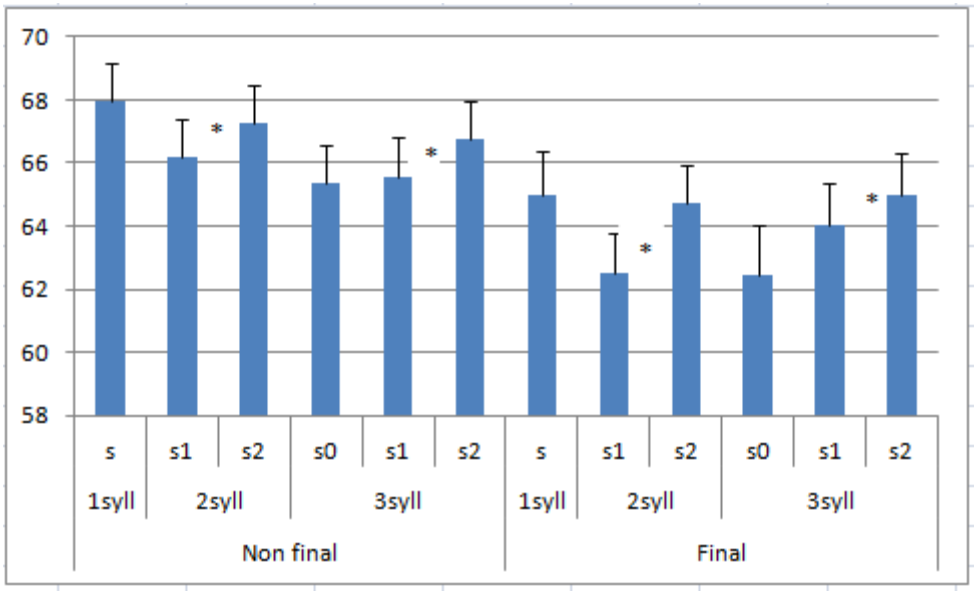


Figure 2: Mean intensity (dB). s, s0, s1 and s2 are syllable positions in segmented chunks. 1syll, 2syll, and 3syll: number of syllables in chunks. Non final and final are positions of chunks in utterances. The symbol * means p < 01.

2.7. Discussion and conclusion

In summary the duration and intensity results mirror each other in indicating an iambic pattern (s2>s1) of acoustic prominence of bisyllabic words/phrases/chunks in narrative speech, consistent with Nguyen and Ingram (2007a, b) that in a bisyllabic unit of speech, the second element is more acoustically prominent than the first element. Generally, the patterns of acoustic results are consistent with the researcher's segmentation in indicating clear units of one–syllable, two-syllable or three-syllable chunks, particularly the final syllable of the speech unit always had the most acoustic prominence. The results mirror those found in Nguyen (2010) and Nguyen (2013) that polysyllabic units of speech in Vietnamese tend to be parsed into bi-syllabic (s2>s1) or tri-syllabic (s2>s0~s1) iambic feet and when syllables stood by themselves, they tended to be lengthened to fill the bi-syllabic foot template. This result also reflects observations by Thompson (1965) and Jones and Huỳnh (1960) that "fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress". The results also support Schiering, Bickel and Hildebrandt (2010) remarks that "Vietnamese provides ample evidence for a genuine stress domain that is preferably disyllabic and maximally trisyllable. Within this domain, stress is realised on the final syllable in the default case. Crucially, this domain is computed irrespective of the morphosyntactic status of its constituent syllables, i.e. stress phonology does not distinguish between a word-level and a phrasal-level of prosodic structure. Metrically, polysyllabic words are thus indistinguishable from other combinations of syllables. Since the most complex structures which are referenced by the rules for iambic rhythm are phrasal, stress may most adequately be attributed to the prosodic domain of the Phonological Phrase."(p.673). It is wondered whether the results of this study is extended to spontaneous speech which needs to be investigated in future studies.

References

Boersma P. & Weenink D. (2007). Praat: doing phonetics by computer (Version 4.5.18) [Computer program]. Retrieved March 30, 2007, from http://www.praat.org/

Cao, X. H. (2003).*Vietnamese: issues of phonology, grammar and semantics*. Education press.

Chaudhary, C. C. (1983). Word stress in Vietnamese: A preliminary investigation. *Indian Linguistics* 44:1-10.

Do, T. D. (1986). *Elements pour une e'tude comparative de l' intonation en Francais et en vietnamien: L'accent de mots en vietnamien*. Memoire de DEA (Universite' de Paries 3 ILPGA.

Durand, M. Maurice, and Nguyen, Tran Huan (1985). *An Introduction to Vietnamese Literature*, translated from the French by D.M. Hawke. New York: Columbia University Press.

Gsell, R. (1980). Remarques sur la structure de l' espace tonal en Vietnamien du sud (Parler de Saigon). *Cahiers d'etudes Vietnamiennes*, 4 (Universite' Paris 7).

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds.), *Phonetics and Phonology, Vol 1: Rhythm and Meter*. San Diego: Academic Press. pp. 201-260.

Hoàng, T.and Hoàng, M. (1975). Remarques sur la structure phonologique du Vietnamien. *Essais Linguistiques*. Etudes Vietnamiennes 40.

Jones, R. B. and Huỳnh, S. T. (1960). *Introduction to Spoken Vietnamese*. Washington, D.C

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications

Nguyễn, D. L. (1970). *A contrastive phonological analysis of English and Vietnamese* (Pacific Linguistics Series, No 8). Canberra: Australian National University.

Nguyễn, T.A.T. and Ingram J.C. (2007a). Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America*, 112 (3), 1746-1757.

Nguyễn, T.A.T. and Ingram J.C. (2007b). Stress and tone Sandhi in Vietnamese reduplications. *Mon-Khmer Studies*, 37, 15-40.

Nguyễn, T.A. T. (2010). Rhythmic pattern and corrective focus in Vietnamese polysyllabic words. *Mon-Khmer Studies*, 39, 1-28.

Nguyễn, T.A. T. (2013). Acoustic and perceptual correlates of Vietnamese folk poetry rhythmic structure. *Journal of the Southeast Asian Linguistic Society*, 6, 54-77.

Schiering, R., B. Bickel, & K. Hildebrandt. (2010). "The prosodic word is not universal, but emergent*," Journal of Linguistics* 46, 657-709

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed), *Nordic Prosody II*, Trondheim: TAPIR.

Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and M.-L. Kean (eds), *Juncture*, Anna Libri, PO Box 876, Saratoga, Calif. 107-129

Thompson, L. (1965). *A Vietnamese Reference Grammar*. Honolulu: University of Hawaii Press.

Appendix 1: the first excerpt of a short story

TÔI ĐI HỌC, a short story by Thanh Tịnh

Hằng năm cứ vào cuối thu, lá ngoài đường rụng nhiều và trên không có những đám mây bàng bạc, lòng tôi lại nao nức những kỷ niệm hoang mang của buổi tựu trường.

Tôi không thể nào quên được những cảm giác trong sáng ấy nảy nở trong lòng tôi như mấy cành hoa tươi mỉm cười giữa bầu trời quang đãng.

Những ý tưởng ấy tôi chưa lần nào ghi lên giấy, vì hồi ấy tôi không biết ghi và ngày nay tôi không nhớ hết. Nhưng mỗi lần thấy mấy em nhỏ rụt rè núp dưới nón mẹ lần đầu tiên đến trường, lòng tôi lại tưng bừng rộn rã.

Buổi sáng mai hôm ấy, một buổi mai đầy sương thu và gió lạnh. Mẹ tôi âu yếm nắm tay tôi dẫn đi trên con đường làng dài và hẹp. Con đường này tôi đã quen đi lại lắm lần, nhưng lần này tự nhiên tôi thấy lạ. Cảnh vật chung quanh tôi đều thay đổi, vì chính lòng tôi đang có sự thay đổi lớn: Hôm nay tôi đi học.

Tôi không lội qua sông thả diều như thằng Quý và không ra đồng nô hò như thằng Sơn nữa.

Trong chiếc áo vải dù đen dài tôi cảm thấy mình trang trọng và đứng đắn.

Dọc đường tôi thấy mấy cậu nhỏ trạc bằng tôi, áo quần tươm tất, nhí nhảnh gọi tên nhau hay trao sách vở cho nhau xem mà tôi thèm. Hai quyển vở mới đang ở trên tay tôi đã bắt đầu thấy nặng. Tôi bặm tay ghì thật chặt, nhưng một quyển vở cũng chì ra và chênh đầu chúi xuống đất. Tôi xóc

lên và nắm lại cẩn thận. Mấy cậu đi trước o sách vở thiệt nhiều lại kèm cả bút thước nữa. Nhưng mấy cậu không để lộ vẻ khó khăn gì hết.

Tôi muốn thử sức mình nên nhìn mẹ tôi:

- Mẹ đưa bút thước cho con cầm.

Mẹ tôi cúi đầu nhìn tôi với cặp mắt thật âu yếm:

- Thôi để mẹ nắm cũng được.

Tôi có ngay cái ý kiến vừa non nớt vừa ngây thơ này: chắc chỉ người thạo mới cầm nổi bút thước.

Appendix 2

An example of the segmented chunks

Tôi/ không thể nào/ quên được/ những cảm giác/ trong sáng/ ấy /nảy nở/ trong lòng tôi/ như/ mấy cành/ hoa tươi /mỉm cười/ giữa bầu trời/ quang đãng/.