博士論文

Mandarin Chinese Tone 3 Sandhi as a Prosodic Cue in Lexical Processing

（語彙処理過程における中国語三声変調の役割）

陳　姿因

Abstract

This thesis investigates how native listeners use Mandarin tone 3 sandhi (T3S) as a prosodic cue in online lexical processing. Previous research shows that listeners utilize prosodic cues to guide lexical analysis. Hirose and Mazuka (2015, 2017) demonstrated that people make use of the accentual change derived from the Tokyo Japanese compound accent rule (CAR; see Kubozono, 2008, for a review) to anticipate lexical structure. For example, when the lexically determined accent word *mikan* ‘tangerine’ (high-low-low tone) is pronounced without the accent (low-high-high tone), listeners expect the accentually changed *mikan* to be the modifier of a compound noun such as *mikan-ju:su* ‘tangerine juice’ (low-high-high + high-low-low tone) rather than an independent single noun.

T3S is a phenomenon where a T3 syllable becomes a T2 syllable when it precedes another T3 syllable (e.g., *ni3* ‘you’ [你] + *hau3* ‘good’ [好] → *ni2* *hau3* ‘hello’). In other words, the necessary conditions of T3S are that a T3S syllable must be followed by another syllable and that the following syllable must be of the T3 type. Two visual world paradigm (VWP) experiments were employed to address the main questions of this dissertation: (i) whether a T3S syllable helps to anticipate lexical structure, leading listeners to expect a compound noun structure with a following syllable over a single noun alternative; and (ii) whether a T3S syllable helps to predict the tone type of the following component (i.e., by activating only candidates of the T3 type).

The first VWP experiment (Chapter 5) found that when hearing input including a T3S syllable (e.g., *zhu2 sun3* ‘bamboo-shoot’ [竹筍] → *zhu2 sun2*…), listeners sent more looks to compound objects (e.g., a picture of a bamboo-shoot and cow combination; *zhu2 sun3* + *ru3 niu2* [竹筍 + 乳牛]) and fewer looks to single objects (e.g., a picture of a bamboo-shoot; *zhu2 sun3* [竹筍]) than they did in baseline conditions. Furthermore, the effect appeared before the listeners encountered the head noun. These results suggest that T3S information helps listeners anticipate lexical structure and that the anticipation relies solely on the tonal information itself, not the segmental information of the coming word.

The second VWP experiment (Chapter 6) found no predictive effect of a T3S syllable on upcoming tone type. In an auditory condition with input including a T3S modifier (e.g., *zhu2 sun3* ‘bamboo-shoot’ → *zhu2 sun2*…), no increase of looks was observed to visual candidates with T3-type head nouns (e.g., a picture of a bamboo-shoot and cow combination; *zhu2 sun3* + *ru3* *niu2* [竹筍 + 乳牛]) compared to looks to non-T3-type head nouns (e.g., a picture of a bamboo-shoot and butterfly combination; *zhu2 sun3* + *hu2* *die2* [竹筍 + 蝴蝶]). These results suggest that T3S does not help listeners predict the

phonological form of the following component. Together, the results of Experiment 1 and Experiment 2 indicate that listeners treat the computation of lexical structural and the retrieval of lexically stored acoustic information as different types of processes.

The reasons for using T3S in this investigation of lexical processing are two. First, Mandarin T3S can be used to examine predictive effects on the phonological form of an upcoming component while rules like the Japanese CAR cannot (because the CAR applies to a compound word regardless of the original accentual pattern of its head). Second, T3S application does not entail a one-to-one association with its morphosyntactic or prosodic environment (cf. the CAR, which obligatorily applies to every compound structure). That is, T3S application is flexible (i.e., a T3S syllable can precede different types of boundary in different positions) and optional (i.e., an underlying T3 syllable in a T3S environment is allowed to remain unchanged). The lack of one-to-one association causes processing ambiguities, and listeners may avoid using such associations as a cue. However, the study's results contradict this possibility by showing that Mandarin listeners are capable of using T3S as a prosodic cue in lexical processing.

Table of Contents

Acknowledgements

I cannot imagine how I would have completed this dissertation without the help of numerous people. First and foremost, I would like to thank my supervisors, Takane Ito and Yuki Hirose. They treated me as a daughter or sister, gave me the strictest training and the warmest support in all respects, and, sometimes, endured my obstinacy. They have been true role models for me in the last ten years, and will be so for the rest of my life.

I would like to thank the other three members of my committee, Shin-ichi Tanaka, Masayuki Yoshikawa, and Hiromu Sakai, for being so careful and thorough with their comments. These valuable comments kicked me out of a rut in my thinking at the last minute.

I am grateful to Manabu Arai, Chie Nakamura, Douglas Roland, Franklin Chang, and Nobuyuki Jincho for teaching me in detail about the techniques of eye-tracking measurement and statistical methods. I hope I can do as well for others in the future. I also thank Laurie Durand for excellent English editing. She helped my chapters change from heavy chanting to a lively sonata.

The faculty and graduate students in the Department of Language and Information Sciences at The University of Tokyo have contributed a great deal to my research life. Sincere thanks go to Kairong Yang, Hideki Ono, Ichiro Kanamaru, Daisuke Miyamoto, Natsumi Shibata (Honda), Jun Yashima, Guanxian Li, Pamela Hsiaowen Peng, Jialiang Li, Fumiaki Nishihara, Chunhua Bai, Toshiyuki Yamada, Shodai Uchida, Chuyu Huang, Itsuki Minemi, Saki Tsumura, Takeshi Kishiyama, Kei Furukawa, and Masataka Ogawa, for creating a great community and for countless inspiring discussions.

Japan Rotary Club, Japan-Taiwan Exchange Association, RIKEN, and The University of Tokyo provided me financial security through scholarships and career opportunities. I am also grateful for their kind understanding when I could only concentrate on my own work.

Special thanks go to Hui-mian Lee, my supervisor at National Taiwan University during my undergraduate degree, who cleared the research road to Japan for me. I also thank my friends at the Odakyu Sightseeing Service Center, Shinjuku Bus Terminal, and Kyoritsu Dormitory, for their encouragement and continually reminding me that there can be more to life than research.

Finally and most importantly, I dedicate this dissertation to my family: Yang-chun Chen, Mei-fan Hung, Yun-su Chen, and Yi-ho Chen. Thank you for your deep love and always waiting for my phone calls on the other side of the sea.

List of Abbreviations

Acc	accusative case
Adj	adjective
Adv	adverb
AP	adjective phrase
Aux	auxiliary verb
CAR	the compound accent rule of Tokyo Japanese
Dat	dative case
HLL	high-low-low tone
lexical T3 syllable	a syllable of the lexically specified T3 type
Nom	nominative case
Non-T3	non-tone-3-type (= tone 1, tone 2, or tone 4)
NP	noun phrase
PRED	predicate
PP	prepositional phrase
S	sentence
surface T2	surface tone 2 type
surface T3	surface tone 3 type
T1/T2/T3/T4	tone 1/tone 2/tone 3/tone 4
T3S	Mandarin Chinese tone 3 sandhi
T3S rule	a process of tonal change in which a T3 syllable becomes a T2 syllable when it precedes another T3 syllable
T3S application	the process of applying the T3S rule
T3S environment	an environment where an underlying T3 syllable is followed by at least one underlying T3 syllable
T3S syllable	the output when T3S is applied to an underlying T3 syllable
VP	verb phrase
VWP	visual world paradigm
unchanged T3 syllable	a T3 syllable that does not changed to T2 in a T3S environment
underlying T3 syllable	The underlying tone type of what is lexically determined as a T3 morpheme, in contrast to the morpheme’s tone status as surface output

Chapter 1

Introduction

1.1 Introduction

The main goal of this dissertation is to examine what effects, if any, rule-driven tonal changes have on lexical processing. More specifically, the dissertation will examine whether Mandarin Chinese tone 3 sandhi (T3S), the application of which does not entail a one-to-one association with its morphosyntactic or phonological environment, is used as a prosodic cue to predict (i) morphosyntactic structure and (ii) phonological form. Lexical processing involves a variety of forms of information, including phonology, prosody (e.g., tone, stress, rhythm, intonation), morphology, and meaning (Traxler and Gernsbacher, 2006 for a review). The main goals of this dissertation are related to the interaction between prosodic processing and morphological processing at the lexical level.

1.2 One-to-one association used as a processing cue

Every utterance carries a considerable amount of information; however, most listeners can process all of this information without delay. It is reasonable for listeners to employ as much information as they can to process current input and predict upcoming materials. Psycholinguistic research has found abundant evidence that if there is a one-to-one association between different elements of linguistic information, people use that association as a cue in anticipatory processing.

For example, in English, vowel duration decreases when the vowel is embedded in a longer word (e.g., /æ/ in *capture* is of shorter duration than /æ/ in *cap*). Blazej and Cohen-Goldberg (2015) demonstrated that vowel duration plays a role in the segmentation of connected speech into lexical units. That is, when listeners are forced to choose between a monomorphemic word (e.g., *clue*) and a multimorphemic word (e.g., *clueless*), they can use vowel duration as a cue to distinguish suffixed words (the short /klu/ of *clueless*) from unsuffixed ones (the long /klu/ of *clue*) before the phonemic differences are encountered (i.e., at the stage of hearing the identical onset of *clue*...).

Hirose and Mazuka (2015, 2017) discussed another example of one-to-one association used as a processing cue: the compound accent rule (CAR) of Tokyo Japanese (Kubozono, 2008 for review). As we will review in detail in Section 1.5, a lexically accented word in Tokyo Japanese loses its accent when compounded with a head noun (e.g., the single word *mikan* 'tangerine' has high-low-low [HLL] tone while the compound word *mikan-ju:su* 'tangerine juice' has low-high-high + high-low-low [LHH+HLL] tone). Because CAR occurs only in compound structures, when an accented word is pronounced without an accent (e.g., *mikan* as LHH), listeners rule out the single word interpretation (e.g., 'tangerine') and relate the change of accentual pattern to a compound interpretation (e.g., a compound word modified by *mikan*).

These studies suggest that vowel duration and accentual change are good cues in predictive processing. In contrast, little is known about whether tonal change has a similar predictive effect. Tone sandhi is a prosodic change that is often observed in tonal languages, where the prescribed tonal pattern on an individual word or morpheme changes according to its morphosyntactic or prosodic environment. If tone sandhi does have a predictive effect like vowel duration or accentual change, the morphosyntactic or prosodic environment related to tone sandhi should be predictable when listeners encounter the output of tone sandhi. However, there are very few studies investigating how tonal change is utilized in online prediction. For this reason, this dissertation study investigates a representative example of tone sandhi, Mandarin Chinese tone 3 sandhi (T3S).

When a certain phenomenon only occurs and must occur in a certain environment (e.g., CAR only applies and must apply to compound structures), we describe the phenomenon as forming a one-to-one association with its environment. The studies mentioned above also suggest that if different elements of linguistic information have a one-to-one association, the association can be used as a processing cue to predict the upcoming materials. However, it is unclear whether any predictive effect occurs in the case that an association between different elements of linguistic information is not a one-to-one relationship. A non-one-to-one association causes processing ambiguities. These ambiguities lead to the risk of wrong predictions, and listeners may not use such high-risk cues. As we will review in Section 1.4, T3S does not have a one-to-one association between its application and its morphosyntactic or phonological environment. We will therefore use T3S to examine whether a non-one-to-one association of linguistic information can be used as a cue in predictive processing.

1.3 Mandarin tone 3 sandhi (T3S)

Mandarin Chinese is a tonal language with a four-tone system. Every word or morpheme is assigned a prescribed pitch pattern (tonal pattern). The association between the word (morpheme) and its tone is lexically determined (e.g., *ni2*[1] ‘mud’ [泥], *ni3* ‘you’ [你]). However, the lexically determined tone may change in particular morphosyntactic or prosodic environments. For example, when an underlying tone 3 (T3) syllable is followed by another underlying T3 syllable, the prior T3 syllable changes to surface tone 2 (T2; e.g., *ni3* ‘you’ [你] + *hao3* ‘good’ [好] → <u>*ni2*</u> *hao3* ‘hello’). This is the phenomenon known as Mandarin Chinese tone 3 sandhi (T3S), which can occur when two conditions are met: one is that the T3S syllable must be followed by another syllable, and the other is that the following syllable must be of the T3 type. The phenomenon of T3S has been a focus of heated discussion in research on Chinese prosody (for reviews, see Chen 2000; Duanmu 2007).

This dissertation asks two specific research questions about the predictive effects of tone sandhi:

(1)
Research Question 1: Does T3S help listeners to predict morphosyntactic structure?
Research Question 2: Does T3S help listeners to predict phonological form?

1.4 Three uncertainties concerning T3S

If Mandarin T3S were a one-on-one association as Japanese CAR is, it would be reasonable to expect that T3S has a predictive effect concerning information in the following syllable. However, empirical data show variety in the application of T3S, indicating that it does not entail a one-to-one association with a morphosyntactic and/or phonological environment. This absence of a one-to-one association mainly results from two uncertainties regarding T3S application: its flexibility and its optionality. In addition, another uncertainty results from the ambiguity in whether input of rising tone should be mapped onto a T2 morpheme or a T3 morpheme. The following subsections discuss these three uncertainties.

1.4.1 The ambiguity of the mapping of rising tone

When the T3S rule applies to an underlying T3 syllable (“underlying T3” refers to the underlying tone type of what is lexically determined as a T3 morpheme, in contrast to

---

[1] The number that follows a syllable without brackets such as *ni3* represents the tone type based on the Mandarin Chinese four-tone system; *ni1* is tone 1 (T1), *ni2* is tone 2 (T2), *ni3* is tone 3 (T3), and *ni4* is tone 4 (T4). The acoustic characteristics is described in Section 3.1.1.

the morpheme's tone status as surface output), the syllable changes to surface tone 2 (T2). For this reason, surface T2 can be associated with either an underlying T2 morpheme or an underlying T3 morpheme. In psycholinguistic terms, without an evident semantic restriction, an input of surface T2 should be ambiguous for listeners in regard to whether it is, underlyingly, a T2 morpheme or a T3 morpheme.

Logically, there are two necessary conditions for recognizing a T3S syllable ("T3S syllable" refers to the output when T3S is applied to an underlying T3 syllable). The first is recognizing the rising tone (surface T2), and the second is mapping the rising tone onto a T3 morpheme. The possible predictive effects of T3S that this dissertation investigates are based on the assumption that an input of rising tone can first be correctly recognized as a T3S syllable. However, there is potential interference when the ambiguity of rising tone cannot be resolved. That is, if there is not sufficient information to require the surface T2 input to be linked to a T3 morpheme, the surface T2 syllable may not be correctly recognized as a T3S syllable. If listeners cannot definitively recognize a T3S syllable (i.e., they cannot exclude the possibility of its being a lexical T2 syllable), the predictive process that should be triggered by T3S may not occur. Hence, the ambiguity in the mapping of a rising tone is a potential factor to weaken the predictive effect of T3S.

The ambiguity that occurs when a surface prosodic form can be mapped onto different morphemes is observed not only in T3S but also in various phenomena in other languages. For example, Japanese *mikan* 'tangerine' [蜜柑] has lexical HLL tone while *mikan* 'not finished' [未完] has lexical LHH tone. When they are compounded with a head *repo:to* 'report' [レポート] as *mikan-repo:to*, they both have LHH tone. Thus, the interpretation for listeners of an input of *mikan* with LHH tone could be ambiguous between 'tangerine report' (e.g., a report concerning tangerine production) and 'unfinished report'.

1.4.2 The flexibility of T3S application

The basic T3S rule can be formalized as T3→T2/_T3. Moreover, the T3S rule is recursive: if there is a sequence of underlying T3 syllables, the rule can maximally apply to every underlying T3 syllable except the last one. (Actually, the T3S rule may NOT apply to every underlying T3 syllable in some cases; Section 1.4.3 describes the optionality of T3S application.) In extreme cases, T3S can apply to strings of T3 syllables that cross a phrasal boundary (2a) or a sentential boundary (2b):

(2)

a. T3S sequence crossing a phrasal boundary

[[Yu3 san3]$_{NP}$ [hen3]$_{Adv}$ [xiao3]$_{AP}$]$_{S}$
rain umbrella very small
→ 2 2 2 3
‘The umbrella is small.’
雨傘很小。

b. T3S sequence crossing a sentential boundary

[[Yu3 san3]$_{NP}$ [[hen3]$_{Adv}$ [[xiao3]$_{A}$]$_{AP}$]$_{S}$ / [[wo3]$_{NP}$ [dei3]$_{Aux}$ [zi4 ji3]$_{PP}$ [cheng1]$_{VP}$]$_{S}$
rain umbrella very small I have to self hold
→ 2 2 2 2/ 2 3 4 3 1
‘The umbrella is small./ I have to hold it by myself.’
雨傘很小。我得自己撐。

In (2a), T3S applies to the noun phrase *yu3 san3* and the adjective *hen3*. In (2b), T3S applies to every syllable in the first sentence and also the first syllable (i.e., the subject) of the second sentence (*yu3 san3 hen3 xiao3/wo3*). In psycholinguistic terms, if we could confirm that T3S cannot apply across a sentence boundary, the occurrence of a T3S syllable such as *hen3→2* ‘very’ in (2a) could help listeners predict a lexical T3 syllable to follow (“lexical T3” refers to the output of an underlying T3 syllable, which retains T3 tone on the surface because it is not followed by another underlying T3 syllable), and more importantly, a sentence boundary to follow the last lexical T3 syllable. However, the empirical data show that we cannot in fact confirm this. As shown in (2b), the T3S rule can apply across a sentence boundary. Therefore, the occurrence of a T3S syllable is not a reliable cue to predict the position of the following sentence boundary.

The T3S rule can apply not only across a sentence boundary but also across other levels of boundaries. For example, underlying T3 in *yu3* ‘rain’ can change to surface T2 both in a compound noun such as [*yu3 / san3*]$_{N}$ ‘rain umbrella’ and in a sentence such as [[*yu3*]$_{NP}$[*xiao3*]$_{AP}$]$_{S}$ ‘rain is small’. That is, the T3S rule can apply across the morphemic boundary in [*yu3 / san3*]$_{N}$ and across the phrasal boundary in [[*yu3*]$_{NP}$[*xiao3*]$_{AP}$]$_{S}$. In psycholinguistic terms, the input of the T3S syllable *yu2* ‘rain’ can be interpreted as a part of a compound noun [*yu3→2/...*]$_{N}$ or as a part of an independent noun phrase [[*yu3→2*]$_{NP}$…]$_{S}$ that will be followed by a predicate. Because the occurrence of T3S is not strongly associated with a specific structure, it may not be a reliable cue to predict what level of boundary (morphemic or phrasal) will come after. In other words, the input of a T3S syllable can help listeners predict a syllable to follow (the first necessary condition; Section 1.2), but it is not known whether listeners are able to relate the following syllable to a specific lexical structure.

1.4.3 The optionality of T3S application

The third potential cause of interference is the optionality of T3S application: it is possible for T3S to remain unapplied even in an environment that satisfies the two requirements of T3S application (hereafter, a "T3S environment" meaning specifically an environment where an underlying T3 syllable is followed by at least one underlying T3 syllable). For example, a word string including two underlying T3 syllables such as *zhu2* ***sun3** + huo3* *guo1* 'bamboo-shoot hot-pot' [竹筍火鍋] can be pronounced in both ways in native uses: *zhu2* ***sun2** huo3* *guo1* or *zhu2* ***sun3** huo3* *guo1*. That is, *sun3* in a T3S environment is allowed to remain unchanged *sun3* (henceforth, a T3 syllable that is not changed in a T3S environment will be termed an "unchanged T3 syllable"). Chapter 3 discusses this point in greater detail, showing that the presence of an unchanged T3 syllable is not random but is controlled by a set of factors such as lexical/syntactic structure and speech speed, by which an unchanged T3 syllable occurs readily as the rightmost syllable of a foot/morphemic/phrase boundary in slow speech.

From a psycholinguistic point of view, as mentioned in Section 1.2, it would be reasonable if T3S could be used to predict the tone type of the upcoming syllable: the occurrence of a T3S syllable should help listeners predict that the following syllable will be of the underlying T3 type, and by contrast, the occurrence of a surface T3 syllable should help listeners predict that the following syllable will be of a non-T3 type. Namely, the tone types of the syllables following T3S and surface T3 input should be in complementary distribution (T3S [surface T2] →T3 vs. surface T3 → non-T3). However, as shown in empirical data, T3S is allowed to remain unapplied in a T3S environment. Therefore, the association of surface T3 → T3 is allowed as well. As a result, there may be no restriction on the tone type that comes after a surface T3 syllable (surface T3 → non-T3 or T3), which may weaken the predictive effect of T3S. In other words, a surface T3 syllable may lose its predictive effect regarding the tone type of the following syllable because surface T3 can be followed by either T3 or non-T3.

1.5 The difference between CAR and T3S

To allow a comparison of a one-to-one association and a non-one-to-one association in terms of their predictive effects, this section provides a basic explanation of the Japanese CAR as a representative one-to-one association. Comparing Japanese CAR and Mandarin T3S suggests possible differences in the processing of lexical structure and phonological form.

1.5.1 The compound accent rule of Tokyo Japanese (CAR)

In Japanese, accentuation is lexically determined (e.g., *mikan* 'tangerine' is initially accented [HLL] while *ringo* 'apple' is unaccented [LHH]). When a morphologically single noun is compounded with a long noun (usually trimoraic or longer), the accent of the compound is obligatorily reassigned to the head noun, and at the same time, the original accent of the modifier noun is overridden (see Kubozono, 2008 for a review). Taking the compound formation of *mikan/ringo-ju:su* 'tangerine/apple-juice' for an example, the juncture accent is reassigned to the head noun *-ju:su* and the original accent on the modifier *mikan/ringo* is overridden. CAR in fact applies to both *mikan-ju:su a*nd *ringo-ju:su* ut it results in two patterns of accentual process. When the modifier noun is a lexically accented word, the accent on the modifier disappears with compound status (e.g., *mikan*: HLL → *mikan-ju:su*: LHH-HLL); in contrast, when the modifier noun is originally unaccented, the tonal pattern of the modifier does not change in the compound (e.g., *ringo*: LHH → *ringo-ju:su*: LHH-HLL).

In comparing CAR with T3S, three characteristics of CAR should be noted. First, CAR only applies to compound structures. Second, CAR application is obligatory. These two characteristics of CAR form a one-to-one association between prosodic change and morphological structure. That is, any accentual change caused by CAR only occurs in compound structures. If listeners use the one-to-one association of CAR in online lexical processing, they should analyze current input of a noun with accentual change as the modifier of a compound and expect the upcoming component to be the head of the compound. The third characteristic is that CAR applies irrespective of the accent patterns of the components in a compound. That is, the accent pattern of the head is irrelevant in CAR application and hence the accentual change on the modifier cannot help the listener to predict the accentual type of the following head. The following subsections introduce three characteristics of T3S that differentiate from CAR.

1.5.2 Predictive effect on lexical structure

Hirose and Mazuka (2015, 2017) found that when a lexically accented word (e.g., *mikan* 'tangerine') changes to LHH, listeners eliminate a single word interpretation and anticipate a compound structure. Furthermore, when a lexically accented word remains unchanged (e.g., *mikan*, HLL), listeners immediately eliminate the possibility of a compound word and retain the single word reading. Thus, in the case of CAR, both its application and the lack of its application are informative.

If T3S had a one-to-one association with a lexical structure, in an environment including only compound structures (e.g., *zhu2* ***sun3*** *+ huo3 guo1* ‘bamboo-shoot hot-pot’ [竹筍火鍋]) and single-word structures (e.g., *zhu2* ***sun3*** ‘bamboo-shoot’ [竹筍]), a T3S syllable would help listeners predict a compound structure. However, as mentioned in Section 1.4.2, the flexibility of T3S may interfere with the predictive effect of T3S in regard to lexical structure. That is, the flexibility of T3S application allows a T3S syllable to be in an independent noun phrase (e.g., [[*zhu2* ***sun3*** →*2*]$_{NP}$…]$_{S}$ ‘bamboo-shoot’ [竹筍] as well as in the modifier of a compound noun (e.g., [*zhu2* ***sun3*** →*2/...*]$_{N}$) in native use. Therefore, the occurrence of a T3S syllable may not help listeners predict a compound structure. If the flexibility of T3S interferes with the predictive effect of T3S for lexical structure, then T3S application is a high-risk cue because it may lead to wrong predictions. Listeners may not use such a high-risk cue to predict lexical structure.

1.5.3 The constancy of the tonal type of the following syllable

Another point in which T3S differ from CAR concerns whether there are restrictions on the original prosodic pattern of the following component. As discussed above, CAR applies to a compound structure no matter what the original accentual pattern of its head is, and hence the accentual change on the modifier (e.g., *mikan*: LHH-) cannot provide any information for listeners to predict the original accentual pattern of the head noun.

This is not the case with T3S. Because the following syllable of a T3S syllable must be of the T3 type, then T3S input should help listeners predict a T3 syllable to follow. In this sense, T3S is more informative than CAR. It should also be noted that the prediction here concerns the prosodic feature of the subsequent morpheme, which is lexically determined. Hence, T3S can be used to examine the predictability of the information stored in lexical entries, while CAR cannot. That is why T3S offers a hitherto unexamined testing ground for the predictive effects of tonal change concerning phonological form.

This study also considers possible interference from the uncertainties of T3S. As mentioned in Section 1.4.3, a surface T3 syllable is allowed to be followed by a T3 syllable or a non-T3 syllable, due to the optionality of T3S application. For example, *zhu2 sun3* ‘*bamboo-shoot*’ [竹筍] can be followed by ***huo3*** *guo1 “hot-pot”* [火鍋] (T3 syllable) or ***bian4*** *dang1 "lunch-box'* [便當] (non-T3 syllable). This optional application may interfere with the predictive effect of surface T3 (but not T3S input)

1.6 Organization of this dissertation

This dissertation is organized as follows. Chapter 2 reviews the research concerning anticipatory effects and the role of prosodic information in online processing. Chapter 3 describes the basic rule of T3S, and provides more detail on the three uncertainties of T3S application and the possible interference they may cause to predictive processing. In Chapter 4, a production task conducted to confirm the preference for a T3S syllable in noun-noun compounds is presented. In Chapter 5 and Chapter 6, two visual world paradigm eye-tracking experiments are reported: The first one examines the predictive effect of T3S on lexical structure, while the second one examines the predictive effect of T3S on phonological form. Concluding remarks and suggestions for further research are provided in Chapter 7.

Chapter 2

Prosodic information in predictive processing

As Chapter 1 discussed, taking T3S as an example, the ambiguities caused by a non-one-to-one association may lead to wrong predictions, and listeners may not use such high-risk cues. However, language processing is fast and automatic. In most cases, listeners process what they hear immediately, using incomplete information rather than waiting for the end of a sentence. The language processing system must "guess boldly" to maintain processing efficiency. This chapter first introduces the visual world paradigm, an experimental method often used for investigating predictive processing (2.1). It then describes how prosodic cues are used for processing language information at various levels (2.2). Next, the chapter reviews studies showing that people do not merely integrate new input as they receive it, but also predict upcoming materials (2.3). Lastly, it introduces previous studies that are relevant to this dissertation's experiments (2.4).

2.1 The visual world paradigm (VWP)

We begin by introducing the visual world paradigm (VWP) method for two reasons. First, many of the previous studies to be reviewed in this chapter used this method. Second, the present study also employs the VWP in its two experiments (Experiment 1, Chapter 5 and Experiment 2, Chapter 6). Moreover, this study's analysis is more complicated than that used in most VWP studies, because it compares combinations of auditory conditions and visual objects. Introducing the basic concepts of the VWP in this section will promote better understanding of Chapters 5 and 6's explanation of the analysis procedure for Experiments 1 and 2.

Eye movement is considered to be a good index of the mental representations that listeners construct when performing a cognitive task (see Huettig, Rommers & Meyer, 2010 for a review). A typical visual world paradigm used in comprehension experiments includes auditory stimuli and visual objects. The auditory stimulus describes a situation that may lead to an association with one (or some) of various objects depicted simultaneously on a screen (e.g., the auditory stimulus *The boy will eat the cake* is presented with a visual display containing pictures of several object candidates such as a cake, a ball, and a car in Altmann and Kamide 1999). Research has found that human eyes automatically move to search for plausible target objects (see Findlay & Gilchrist 2003 for a review). People anticipatorily look for corresponding visual objects when they detect information that provides a predictive effect regarding upcoming materials in auditory stimuli (e.g., more looks to a cake picture than to pictures of uneatable candidates when hearing *The boy will eat the…*). In VWP experiments, participants' eye movements are recorded via eye trackers to explore the time course of how the linguistic information included in an auditory condition (or sometimes in a visual condition or a combination of both) is integrated in real-time processing. Because the timing of eye movements can be recorded and measured at a very fine level of detail, this method is sensitive to extremely fine-grained responses to input. It is therefore a very useful tool for investigating the cognitive processing of information from multiple modes (e.g., auditory, visual), and how people predict about-to-be encountered material.

2.2 The use of prosodic information in language processing

It is well known that prosodic information can be used as a cue at different levels of language processing. For example, in a word recognition study, Culter and Norris (1988) found that nonwords with strong first and weak second syllables (e.g., *mintesh*) were easier for listeners to detect than nonwords with two strong syllables (e.g., *mintayve*). The authors explained these findings by suggesting that the occurrence of a strong syllable in English triggers segmentation of continuous speech while a weak syllable does not. This prosodic marking of the onset of each segment provides a cue to start lexical access. Therefore, the nonwords with two strong syllables needed more detection time because they were wrongly analyzed as two components for lexical access.

Vogel and Raimy (2002) tested children's ability to use contrastive prosodic patterns in English as cues in processing lexical structure. They employed sets of pictures of visual objects whose meanings corresponded to compound and phrase interpretations, which differed only by acoustic stress (e.g., the compound *HOT dog* and the phrase *hot DOG*). They found that children older than twelve could use the rhythmic information to distinguish between compounds and phrases (see also Atkinson-King 1973).

Misono, Mazuka, Kondo, and Kiritani (1997) showed that prosodic boundary cues play an important role in syntactic ambiguity resolution in Japanese. Sentence (3) is globally ambiguous; it is grammatical with either a main-clause or embedded-clause interpretation of the adverbial clause *yopparatte* 'being drunk', but the main-clause interpretation of the adverbial clause (3a) would usually be favored according to world knowledge (it is more plausible for a father to get drunk than a baby). Misono et al. found that prosodic boundary cues can override such pragmatic preferences. That is, when the auditory sentence was pronounced with an emphasized boundary after the topic-marked NP ("embedded-clause prosody"), listeners tended to analyze (3) with the embedded-clause interpretation (3b).

(3)

Chichioya-wa	*yopparatte*	*nete-iru*	*akanboo-o*	*ofuro-ni*	*ireyoo to shita.*
NP1-Top	PRED1	PRED2	NP2-Acc	NP3-Goal	PRED3
father	drunken	sleeping	baby	bath into	put

a. Main-clause interpretation: ‘The father, being drunk, tried to bathe the sleeping baby.’
b. Embedded-clause interpretation: ‘The father tried to bathe the baby that was drunk and sleeping.’

As these studies’ results indicate, prosodic information participates in various levels of language processing.

2.3 Anticipatory effects on upcoming material

Numerous studies have found that humans do not only integrate current input with what they have heard so far, but also make predictions regarding upcoming material. For example, in a VWP experiment, Altmann and Kamide (1999) found that the meaning of a verb helps listeners to anticipate plausible objects. When listeners were provided auditory stimuli such as *The boy eats…*, they sent more looks toward an eatable object like a cake in a picture that also contained other, uneatable objects. Furthermore, the effect starts before the direct object NP is auditorily presented. In contrast, a sentence beginning *The boy moves…* does not have the same anticipatory effect. The results suggest that people use the semantics of a verb to anticipate the verb’s object before they encounter the object.

Kamide, Altmann, and Haywood (2003) further found that not only the semantics of verbs, but also types of arguments can help listeners to predict the following materials. Because Japanese is a verb-final language, listeners cannot know the verb until the end of a sentence. For example, in (4a), the dative verb *hakobu* ‘bring’, and in (4b), the accusative verb *karakau* ‘tease’, occur at the end of the sentences.

(4)

a. Dative condition

Weitoresu-ga	*kyaku-ni*	*tanosigeni*	*hanbaagaa-o*	*hakobu.*
NP1-Nom	NP2-Dat	Adv	NP3-Acc	V

waitress customer merrily hamburger bring
'The waitress will merrily bring the hamburger to the customer.'

b. Accusative condition
Weitoresu-ga kyaku-o tanosigeni karakau.
NP1-Nom NP2-Acc Adv V
waitress customer merrily tease
'The waitress will merrily tease the customer.'

However, listeners use the case marking and lexical information of the arguments to anticipate the verb. Kamide et al. showed that listeners cast more looks to the visual object representing a hamburger in condition (4a) and to the visual object representing a customer in condition (4b) before they encountered the verbs. This is because the combination of the nominative NP *weitoresu-ga* and the dative NP *kyaku-ni* in condition (4a) helps listeners anticipate a ditransitive verb, leading to more eye fixations on the hamburger image, which is a suitable argument for a likely ditransitive verb. On the other hand, the combination of the nominative NP *weitoresu-ga* and the accusative NP *kyaku-o* in condition (4b) helps listeners anticipate a transitive verb, leading to more eye fixations on the customer image, which fits with a likely transitive verb.

Ito and Speer's (2008) study provided another example of the anticipatory effect of prosodic information. They found that intonational prominence facilitated the processing of pragmatically contrastive contexts. In English, when an L+H* accent places focus on a subject such as *Katie* in *KATIE did not win a truck*, it is plausible that the sentence will continue with a contrastive reading such that another person did win a truck such as *...LAURA did*. In contrast, when the object is placed in focus by carrying the L+H* accent, as with *truck* in *Katie did not win a TRUCK*, the sentence is expected to continue by naming another thing that Katie actually won such as *...she won a MOTORCYCLE*. In the visual world paradigm utilized in the study, a grid with different shapes of ornaments in different colors (e.g., green/white/red/blue balls in one of cell, and silver/white/ green/orange rings in another cell) was presented as the visual display. The auditory stimuli were two-sentence instructions, which were manipulated by having or not having accentual prominence on an adjective in the second instruction such as, *First, hang the green ball. Now hang the BLUE ball/blue ball*. It was found that the prominent accent on the prenominal adjective (e.g., *BLUE*) evoked a contrastive relation between the immediately preceding referent (e.g., ball in the first instruction) and the about-to-be-mentioned referent (e.g., which should be a ball that is not green but is blue). That is, the contrastive interpretation on the color word caused by accentual prominence helped the listeners predict that the upcoming noun in the second instruction would be the same noun just mentioned in the first instruction. Therefore, earlier fixations on the target cell (the same cell mentioned in the first instruction, e.g., the cell including the four colors of balls) were observed when the participants were hearing the accented adjective in the second instruction. Furthermore, the fixation proportion of the accented adjective condition showed a steeper rise that diverged from the condition without accentual prominence at around 100ms after the onset of the head noun, which indicates that the contrasted context had been constructed before the segmental information of the head noun was fully processed.

These results indicate that listeners not only integrate current input into what they have already parsed, but also have detailed expectations for upcoming materials if there is any association between the input so far and the about-to-be encountered materials.

2.4 Previous studies relevant to the current study

2.4.1 Hirose and Mazuka 2015, 2017: Accent information and lexical structure

As discussed in Section 1.6, this dissertation will examine the predictive effect of T3S on lexical structure in a visual world paradigm experiment, and will then compare the observed effects with the effects of the Japanese CAR, as reported by Hirose and Mazuka (2015, 2017).

The experimental paradigm in this study will follow that used in Hirose and Mazuka's (2015, 2017) studies. Hirose and Mazuka found a predictive effect of prosodic information in lexical processing, in which the accentual information derived from compound prosody in Tokyo Japanese (see Section 1.4) facilitates predictions for lexical structure. They conducted a visual world paradigm experiment to examine how the accentual change on lexically accented modifier nouns helps listeners to anticipate compound structure. The materials consisted of auditory stimuli as in (5) and visual displays as in **Figure 1**.

(5) Auditory stimuli from Hirose and Mazuka 2015, 2017
("C1" refers to the first constituent)

a. accented C1/compound (with accent change on C1)
Mikan-risu wa dore? 'Where is the tangerine-squirrel?'
HLL + HL → LHH-HL

b. accented C1/simplex
Mikan wa dore? ‘Where is the tangerine?’
HLL
c. unaccented C1/compound (without C1 accent change)
Ringo-koara wa dore? ‘Where is the apple-koala?’
LHH + HLL
d. unaccented C1/simplex
Ringo wa dore? ‘Where is the apple?’
LHH

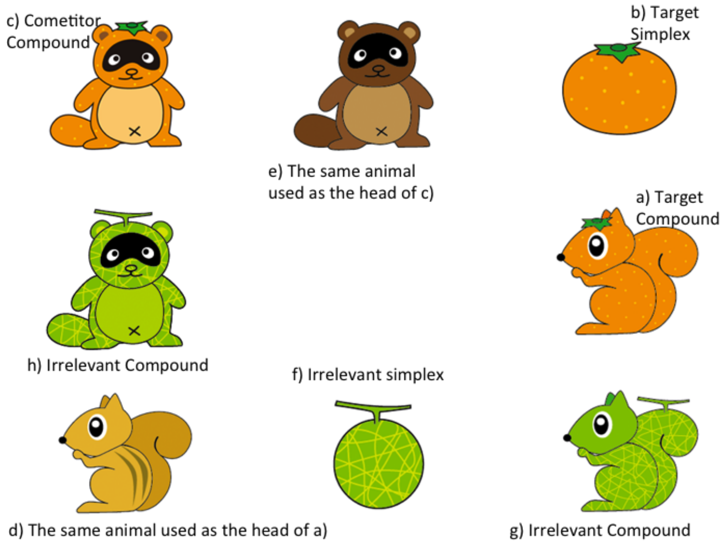


*Figure 1***.** An example of the visual stimuli of Hirose and Mazuka 2015, 2017

The results demonstrate that when a word like *mikan* (LHH) undergoes accent change in the auditory condition (5a), it leads to a significant increase of fixations on the target compound object (e.g., the tangerine-squirrel in **Figure 1**) and the competitor compound object (e.g., the tangerine-raccoon) before the onset of the head noun, and a significant decrease of fixations on the target simplex object (e.g., the tangerine) when compared to the originally accented auditory condition *mikan* (HLL) as in (5b). However, there was no difference in fixation rates between the two auditory conditions without accentual changes, as in (5c) and (5d), *ringo/ringo…* (LHH), before listeners encountered the head noun (*-koara*). These results indicate that when listeners detect the signal that CAR is being applied to the current input, they anticipatorily rule out the possibility of a single noun and instead build a compound structure (e.g., they decide it is not a simple tangerine but something else related to tangerines). The anticipation relies entirely on the accentual change at the modifier, not the segmental information of the head noun. Furthermore, the time point at which listeners distinguished the target compound object from the competitor compound object was earlier when they heard the auditory compound condition with the accent change, such as *mikan-risu* (LHH-HL) in (5a), than when they heard it without accent change, as in *ringo-koara* (LHH-HLL) in (5b). This finding suggests that faster recognition of the head in the CAR condition in (5a) is enabled by the early construction of a compound representation at the previous processing stage. If this effect can be generalized to other contexts, we can expect that a Mandarin listener will also use T3S as a cue to predict lexical structure. The effect will be examined in Experiment 1 (Chapter 5).

2.4.2 Ito, Pickering, and Corley 2018: Predicting phonological form

This study will also investigate the predictive effect of T3S on phonological form. As discussed in Section 1.5.2, the CAR is not useful for testing predictions of phonological form because it applies to a compound structure regardless of the original accentual pattern of its head. In contrast, T3S may be a good cue to predict the tone type of an upcoming syllable, allowing this study to focus on how prosodic information might help listeners to predict the phonological form of an upcoming word. There is a previous study showing that phonological form can be anticipatorily predicted.

Ito, Pickering, and Corley (2018) conducted a visual world paradigm experiment in which participants listened to a sentence in English such as *The tourists expected rain when the sun went behind the…*, and viewed a visual display including (i) a target object corresponding to the highly predictable next word (e.g., *cloud*); (ii) a competitor object whose name is phonologically related to the predictable word (e.g., *clown*); and (iii) a distractor object, for a baseline, which was unrelated to the predictable word (e.g., a globe). (There was another condition, irrelevant here, which included an object related to Japanese translation for L2 participants.)

The results showed that listeners looked more often at the target object from 600ms before the onset of the critical noun compared to the baseline. More interestingly, listeners sent more looks to the competitor object from 500ms before the onset of the critical noun compared to the baseline. These results suggest that listeners can predict the phonological form of an upcoming word when the upcoming word is highly predictable. If this is true in regard to Mandarin T3S, we can expect that Mandarin listeners will be able to use T3S's application as a cue to predict the tone type of the upcoming syllable. This effect will be examined in Experiment 2 (Chapter 6).

This chapter has introduced how prosodic information is used at different levels of processing, the anticipatory effect of predictive processing, and relevant studies concerning the predictive processing of lexical structure and phonological form. The next chapter will give more detailed descriptions of the characteristics and the environment for the occurrence of T3S, and consider possible situations in which T3S may be used as a prosodic cue in lexical processing.

Chapter 3

The characteristics of tone 3 sandhi

Chapter 1 introduced T3S, and explained why its application does not have a one-to-one association with its morphosyntactic or phonological environments. The chapter also briefly discussed the characteristics of T3S that may cause interference in the predictive processing of lexical structure or phonological form when the input is a T3S or surface T3 syllable. Chapter 2, however, reviewed evidence that listeners tend to use as much of various types of information as they can to predict upcoming material, even when the information is incomplete. Chapter 2's discussion suggests the possibility that, despite the risks, people may use non-one-to-one associations as processing cues for predicting upcoming materials. Before we investigate whether or not people do indeed use T3S in this way, this chapter explains in detail why T3S application does not form a one-to-one association with its environment, and the potential interference in predictive processing caused by the uncertainties of T3S application. Section 3.1 describes the basic T3S rule. Section 3.2 presents empirical data demonstrating three uncertainties of T3S application, and considers how these uncertainties might influence predictive processing. Section 3.3 explains the necessary conditions for testing T3S as a cue in lexical processing, which this study will use to create an experimental environment similar to the one employed by Hirose and Mazuka (2015, 2017) to test the CAR, which will allow a comparison of the predictive effects of T3S and the predictive effects of the CAR.

3.1 The basics of Mandarin tone 3 sandhi

3.1.1 The lexical four-tone system of Mandarin Chinese

Mandarin Chinese[2] is a lexical tone language. The tone type is characterized by the continuous height of the fundamental frequency (F0) over a syllable. In Mandarin Chinese, a syllable stands as a morpheme, and the tone type of a syllable is determined

---

[2] The stimulus items used in the experiments in this dissertation were based on Taiwan Mandarin Chinese, which has some lexical differences from Beijing Mandarin Chinese in segmentation and tone. Of the 24 items, three (12.5%) have different pronunciations in Taiwan and Beijing Mandarin. (For the items, see **Appendix B** and **Appendix C**.) The participants were all Taiwan Mandarin Chinese native speakers.

Taiwan Mandarin and Beijing Mandarin speakers may not share identical language knowledge and processing parsers, but the two dialects do share the critical characteristics of interest to this study, and it is conceivable that the predictive effects of T3S could be confirmed in both Taiwan and Beijing Mandarin. However, the results of the current study can be generalized only to Taiwan Mandarin native speakers.

by the morpheme. Moreover, one syllable can stand alone as an independent word in the majority of cases. Therefore, tone type can be used to differentiate the meaning of words that share the same segments. The bottom panels of **Figure 2** (blue lines) show the pitch contours of the four different tone types on the same segment, *guo*: the first tone (T1) is high-high [55], the second tone (T2) is middle-rise [35], the third tone (T3) is low-down-rise [214], and the fourth tone (T4) is high-fall [51].[3] The four tone types on the same syllable correspond to different words: *guo1* 'pan' [鍋], *guo2* 'country' [國], *guo3* 'fruit' [果], and *guo4* 'pass' [過].

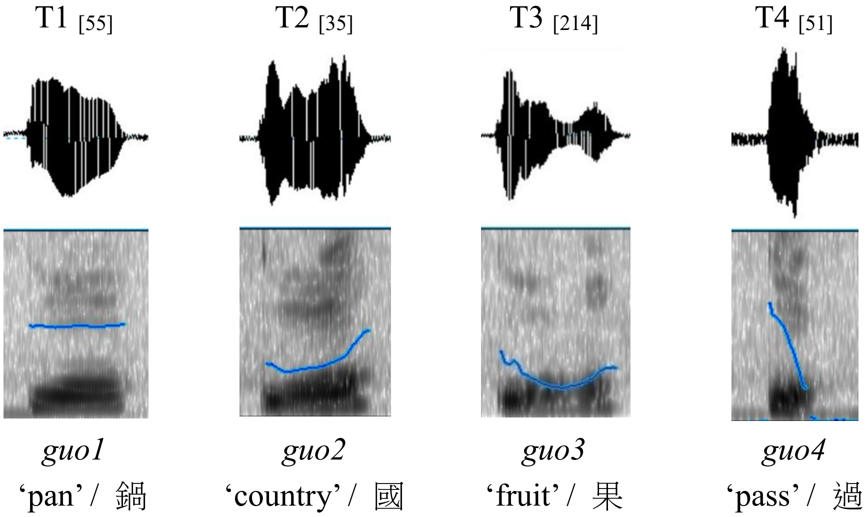


Figure 2. The pitch contours of the four basic tones on the same segment *guo*, which represent different meanings: 'pan', 'country', 'fruit', and 'pass'. The pronunciation was recorded by the author using Praat.

3.1.2 The basics of the T3S rule

Tone sandhi is a phonological process by which lexical tones exhibit contextually determined alternation triggered by the prosodic or morphosyntactic environment surrounding the tone. It is commonly observed in tone languages spoken around the world. In Taiwanese, for instance, every syllable in a T3S environment undergoes tone sandhi

---

[3] The numbers in square brackets describe the relative height of a tone pitch on a five-level scale, where [5] represents the highest pitch and [1] the lowest. The sequences of numbers such as [55] or [35] describe the pitch change within a syllable. This is a traditional notation used in Chinese dialect research since Chao's 1930 study.

except in the final position of a maximal projection. For example, *tai[24] / pak[3] / hue[51] / tshia[44] / tsam[33]* 'Taipei train station' [台北火車站] becomes *tai(33) / pak(5) / hue(44) / tshia(33) / tsam[33]* (Zhang, 2014).[4] Mandarin Chinese tone 3 sandhi (T3S) is also a process of tonal change triggered by adjacent tone, in which a T3 syllable becomes a T2 syllable when it precedes another T3 syllable (see Duanmu, 2007, Ch. 11 for a review). The rule can be formulated as in (6):

(6) T3S rule: T3 → T2/_T3

As shown in **Figure 3**, *shui3* 'water' [水] and *guo3* 'fruit' [果] are both underlying T3 syllables ("underlying T3" refers to a syllable of the lexically specified T3 type, which is contrary to its surface tone type). These words are pronounced with T3 in isolation. However, when they are pronounced in a bisyllabic word in fluent speech, the first underlying T3 syllable *shui3* changes to T2 as in *shui2*.

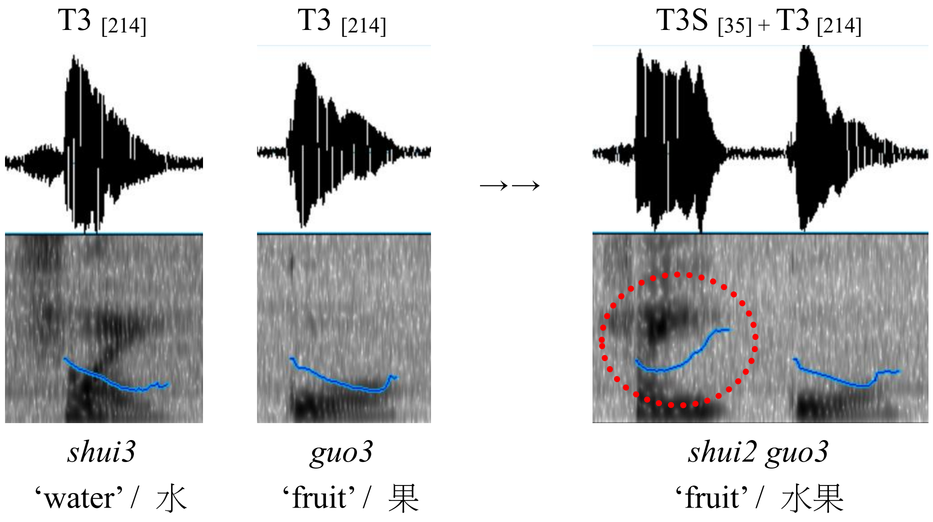


Figure 3. The pitch contours of *shui guo* 'fruit' when pronounced in isolation with its lexical T3 pattern (left) and when T3S is applied to the first syllable (right). The pronunciation was recorded by the author using Praat.

3.1.3 The necessary conditions of T3S application

---

[4] The sequences of numbers in square brackets such as [55] represent lexically determined tonal value on a five-level scale, while those in parentheses such as (33) represent the surface tonal value after tone sandhi.

Cheng (1973) and Yip (2002) proposed that the T3S rule is a case of dissimilation, where two identical tones in sequence, such as T3-T3, tend to be avoided. T3 has a more complicated pitch curve (low-down-rise [214]) than other tone types. It would be reasonable for a T3-T3 syllable sequence with its double curves (low-down-rise + low-down-rise) to be hard to pronounce, leading speakers to avoid it by changing the first T3 syllable to another tone type.

If T3S application is triggered by the need for phonological differentiation, it would be inevitable for every underlying T3 syllable to change to T2 when it is followed by another underlying T3 syllable; in other words, the occurrence of a T3S syllable and a following underlying T3 syllable would form a one-to-one association. Furthermore, if listeners utilized this one-to-one association in predictive processing, an input of a T3S syllable should help listeners to predict an underlying T3 syllable to follow immediately. The condition in which a T3S syllable unambiguously signaled an underlying T3 syllable as a subsequent element could be further subdivided into two subconditions: that a T3S syllable must be followed by another syllable, and that the following syllable must be of the underlying T3 type. The first condition could relate to information on lexical structure and the second condition could relate to phonological information.

3.1.4 The acoustic difference between T3S and T2

In the majority of previous studies, a rising tone that corresponds to a T3S syllable has simply been taken as the same tonal value as that of a lexical T2 syllable (see Chen, 2000; Duanmu, 2007 for reviews). However, some acoustical studies have shown that T3S syllables are physiologically not the same as lexical T2 syllables, raising the question of whether listeners recognize the acoustic differences and map them onto the corresponding surface tone type.

Yuan and Chen (2014) analyzed disyllabic words in a corpus of telephone conversations and broadcast news speech. They found that even though T3S and lexical T2 are both surface rising tones, they have different acoustic properties: the T3S syllable has a smaller amplitude and shorter time span of F_0 rising than the lexical T2 syllable (see also Peng, 2000; Xu, 1997).

It is possible that the acoustic differences play a role in predictive processing. That is, if listeners can differentiate the tone type of T3S and lexical T2 only by their acoustic properties at a stage before lexical perception, the acoustic differences would be an indirect factor that promotes predictive effects once a T3S syllable has been correctly recognized. However, as Section 3.2.1 will show, a rising tone activates both T2 and T3 morphemes even if the context is strongly biased to either a T2 meaning or a T3 meaning. That is, the acoustic differences do not help listeners avoid the temporary ambiguity of a rising tone. Therefore, this study does not consider the influence of the acoustic difference between lexical T2 and T3S syllables on the predictive effects of a T3S syllable.

3.2 Uncertainties concerning T3S application

As mentioned in Section 3.1.3, if T3S application were triggered by a need for phonological dissimilation, it would have a one-to-one association with its morphosyntactic or phonological environment. If this were the case, Mandarin T3S might be easily used as a prosodic cue in predictive processing, as the Japanese CAR is. However, as introduced in Section 1.4, considerable empirical data demonstrate that T3S application does not have a one-to-one association with its morphosyntactic or phonological environment, due to the flexibility and the optionality of T3S application. Furthermore, numerous previous studies have shown the ambiguity of rising tone before a T3S syllable is recognized. The next subsections first introduce relevant studies concerning the ambiguity of rising tone (3.2.1), and then describe in detail the phenomena of the flexibility (3.2.2) and the optionality (3.2.3) of T3S application.

3.2.1 On the ambiguity of rising tone

As **Figure 3** shows, the underlying T3 syllable *shui3* changes to rising tone when it is followed by an underlying T3 syllable *guo3*. The majority of previous studies on Chinese phonology (see Chen, 2000; Duanmu, 2008 for reviews) consider the tone of a T3S syllable to be identical to that of a lexical T2 syllable. This section shows evidence that input of rising tone can be associated with a T2 morpheme (as a lexical T2 syllable) or a T3 morpheme (as a T3S syllable).

At the stage of word recognition, there are two necessary conditions to recognize a T3S syllable. One is perceiving the rising tone (sound), while the other is correctly mapping the rising tone to a T3 morpheme (meaning). Because both lexical T2 and T3S syllables are realized with surface rising tone, if the meaning is not restricted, an input of rising tone may be ambiguous between T2 and T3 morphemes. However, it is impossible for speakers to provide enough contextual restrictions on meaning all the time in speedy conversation. Therefore, in most cases, listeners cannot completely exclude the ambiguity carried by a rising tone in input. For example, input of *guo2* can be taken as a lexical T2 word *guo2* ‘country’ or a T3S output of a lexical T3 word *guo3* ‘fruit’ (as in **Figures 2** and **3**). For this reason, a rising tone cannot be guaranteed to relate to T3 meaning, which may interfere with the predictive effects of T3S.

In addition to the ambiguity of surface T2, Mandarin native speakers may also perceive surface T3 as ambiguous between a T2 morpheme and a T3 morpheme. One possible reason is that surface T2 and surface T3 are reported to be perceptually close. Nevertheless, there is evidence that listeners recognize surface T3 and surface T2 in different ways; namely, surface T2 can activate both a T2 morpheme and a T3 morpheme while surface T3 can only activate a T3 morpheme.

A number of studies have shown that the perceptual distance between surface T2 and surface T3 in nonsense words is very small. Huang (2001) asked participants to judge whether nonsense words with different tones were the same or different. The study found higher error rates and longer reaction times for discrimination between surface T2 and surface T3 (in two orders: T2-T3, 11%, 699ms; T3-T2, 7%, 667ms) than between other pairs of tones (less than 7%, 600ms). Huang used these results as criteria in a two-dimensional scaling and found that surface T2 and surface T3 are perceptually very close compared to the other tones; this closeness may result in confusion in word identification. Chen, Liu, and Kager (2015) also conducted a discrimination task with nonsense words carrying sequences of tones (such as surface T2 + surface T3 or surface T3 + surface T3). They found that listeners could not distinguish T2-T3 from T3-T3 at an above chance level whereas they could easily distinguish other sequences of tones.

These studies indicate that it may be difficult for listeners to differentiate between surface T2 and surface T3, for which there may be two possible explanations. First, because the initial tonal values of surface T2 [35] and surface T3 [214] are both low (compared to T1 [55] and T4 [51]), the perceptual closeness may simply be due to acoustic similarity. Second, the difficulty may be related to overlapping meanings triggered by surface T2 and surface T3. Because lexical T2 and T3S syllables are realized as surface T2, an input of surface T2 may trigger both a T2 meaning and a T3 meaning. As for surface T3, only lexical T3 is realized as surface T3. It is uncertain whether an input of surface T3 can only trigger a T3 meaning or can also trigger a T2 meaning based on perceptual closeness. At the least, however, there is an overlap of T3 meanings between surface T2 and surface T3, which may be what makes it difficult for listeners to distinguish surface T2 from surface T3 in input, especially in the case of nonsense words.

Several studies that have examined the scope of meaning that can be triggered by a homophone (i.e., a syllable related to different meanings) may shed light on the difficulty of differentiating between T2 and T3. Cutler (1986) used an English crossmodal priming task, in which listeners heard bisyllabic words with different stress patterns, such as front stressed *FOREarm* or back stressed *foreARM*, and judged whether meanings such as ‘elbow’ or ‘prepare’ were related to the words. The findings showed that both stress patterns activated both meanings. This suggests that if a syllable has different pitch patterns, the meanings related to every pitch pattern should all be activated simultaneously. If this is true in Mandarin, a four-tone language, then the meanings related to the four tone patterns should all be activated when any syllable is input (e.g., the four meanings in **Figure 2** would all be activated when *guo* with any tone is input). However, the studies conducted by Huang(2001) and Chen et al. (2015) demonstrated that this is not the case, because they show the overlapping of meanings only between surface T2 and surface T3, not among all four tones. This difference may be due to differences in how priming effects work in stress languages (e.g., English) and tone languages (e.g., Mandarin).

Other previous studies have used real words to examine how many meanings can be triggered by surface T2 or surface T3. Shen, Deutsch, and Rayner’s (2013) visual world paradigm experiments found that when a surface T3 sound was the input, listeners could recognize a T3 word in a visual display rapidly and without hesitation. In contrast, when a surface T2 sound was the input, listeners first incorrectly fixated on a T3 word, only moving to the target T2 word after the input word’s offset. These results indicate that surface T3 does not evoke T2 meanings, whereas surface T2 does evoke both T2 meanings and T3 meanings (although at different time points).

Speer and Xu (2008) found a similar tendency as that shown by Shen et al. (2013). They used a crossmodal semantic priming experiment to compare how tonally ambiguous words (i.e., words with contextually restricted meanings and either T3S or T2, which are both rising tones) and unambiguous words (lexical T3, a low-rising tone, with original T3 meaning) are associated with T2 meaning and T3 meaning. Examples of a set of their auditory priming sentences are shown in (7), and of a set of visual targets in (8).

(7) Example of auditory primes used by Speer and Xu (2008)

a. Ambiguous sentence: underlying T2 + lexical T3[5]

Jin1 nian2 hai3 bian1 ***yu2 hen3*** *duo1.*

---

[5] The priming sentences of (7a) and (7b) were auditory stimuli. The surface input of the target word in both is *yu2* (‘fish’ and ‘rain’), with rising tone. Speer and Xu (2008) did not explain in their short abstract how they manipulated the difference in the two ambiguous conditions. It is conjectured that they used context preceding the priming sentence to restrict the meaning in each condition.

this year sea side **fish** **very** much
今年 海邊 **魚** 很 多
‘This year there is a lot of fish near the seashore.’

b. Ambiguous sentence: T3S + lexical T3
Jin1 nian2 *hai3 bian1* ***yu3→2*** ***hen3*** *duo1.*
this year sea side **rain** **very** much
今年 海邊 **雨** 很 多
‘This year there is a lot of rain near the seashore.’

(8) Example of visual character targets used by Speer and Xu (2008)

T2 meaning association:	食品	*shi2 pin3*	‘food’
T3 meaning association:	晴天	*qing2 tian1*	‘sunshine’
Irrelevant:	靈魂	*ling2 hun2*	‘spirit’

The results show that in both the T2 ambiguous sentence (7a) and the T3S ambiguous sentence (7b) conditions, the priming time was shorter for both the T2 meaning association target and the T3 meaning association target. In contrast, in the T3 unambiguous sentence condition,[6] only the T3 meaning association target had a shorter priming time than the other targets. These findings indicate that lexical T2 syllables and T3S syllables (i.e., both with surface rising tone) evoke both T2 meanings and T3 meanings, while a lexical T3 syllable evokes only a T3 meaning.

To sum up the results discussed above, they suggest that surface T2 activates both T2 and T3 meanings, while surface T3 only activates T3 meanings, not T2 meanings. (9) shows the possible patterns for the triggering of meanings by these two surface tones, and also indicates how the meanings can overlap between a T3S syllable and a lexical T3

---

[6] Speer and Xu (2008) did not provide an example of their T3 unambiguous sentence condition. It is conjectured to comprise sentences such as the following:

(i) Unambiguous sentence: Underlying T3 + non-T3
Jin1 nian2 *hai3 bian1* ***yu3*** ***zhen1*** *duo1.*
this year sea side **rain** **really** much
今年 海邊 **雨** 真 多
‘This year there is really a lot of rain near the seashore.’

syllable.

(9) Possible patterns for how meanings are triggered by T2 and T3 surface tones

Surface tone	Underlying tone	Possible meanings: Examples	Syllable type
T2 →	T2 →	魚 'fish', 於 'at'…	Lexical T2
↘	T3 →	雨 'rain', 語 'word'…	T3S
T3 →	T3 →	雨 'rain', 語 'word'…	Lexical T3

Because this ambiguity of semantic mapping only occurs with surface T2, and not with surface T3, we can ignore possible interference from surface T3. However, because the ambiguity of surface T2 occurs before word recognition, which is at a different stage than the phenomenon of interest to this investigation, it is important that the experimental materials exclude the possibility of effects from the ambiguity of rising tone. There are two ways to control this ambiguity. The first way is to use items of T3S syllables without lexical T2 counterparts, or of lexical T2 syllables without T3S syllable counterparts. This would avoid the possibility of activating both T2 and T3 morphemes with input of rising tone. The second way is to provide enough contextual information to force the participant to relate a rising tone to a T3 morpheme. The materials employed in the norming task (Chapter 4) and VWP experiments (Chapters 5 and 6) were created with these considerations in mind.

3.2.2 On the flexibility of T3S application

This section describes the second uncertainty of T3S application, which relates to its flexibility. The T3S rule, repeated here in (6), is recursive: it can be applied to every syllable in an underlying T3 sequence except the last one. Theoretically, for any length of text of underlying T3 syllables that can be uttered before the speaker needs to take a breath, the T3S rule maximally applies to the whole sequence even if it crosses a sentence boundary. Let us again compare the two sequences in (2), repeated here as (10).

(6) T3S rule: T3 → T2/_T3

(10)

a. T3S sequence crossing a phrasal boundary

[[*Yu3* *san3*]$_{\text{NP}}$ [*hen3*]$_{\text{adv}}$ [*xiao3*]$_{\text{AP}}$]$_{\text{S}}$ →$\underline{2\ 2\ 2}$ 3

rain umbrella very small
‘The umbrella is small.’ / 雨傘很小

b. T3S sequence crossing a sentential boundary
[[*Yu3* *san3*]$_{\mathrm{NP}}$ [*hen3*]$_{\mathrm{adv}}$ [*xiao3*]$_{\mathrm{AP}}$]$_{\mathrm{S}}$ / [[*wo3*]$_{\mathrm{NP}}$ [*dei3* [*zi4 ji3*]$_{\mathrm{PP}}$ [*cheng1*]$_{\mathrm{VP}}$]$_{\mathrm{S}}$
rain umbrella very small I have to self hold
→ 2 2 2 2/ 2 3 4 3 1
‘The umbrella is small. / I have to hold it by myself.’ / 雨傘很小。我得自己撐

In (10a) and (10b), all the underlying T3 syllables in the sequences except the last one change to surface T2. If (10a) were the only available data, we might wrongly infer that the T3S rule always maximally applies within a sentence. In this case, input of a T3S syllable would indicate that the sentence was not going to end immediately. The sentence boundary would have to be after the last lexical T3 syllable (…T3S → lexical T3 → sentence boundary). Such an association could help listeners predict what remained to be said before a sentence ended when a T3S syllable was the input.

If we compare (10a) with (10b), however, we see that it is possible for the T3S rule to apply across a sentence boundary. *Xiao3* ‘small’ changes to surface T2 even though it is next to the sentence boundary (T3S → sentence boundary → T3S). That is, the occurrence of a T3S syllable is not a safe signal of an upcoming sentence boundary.

Nor can a T3S syllable be taken as a hint on structural boundaries at other levels. For example, the second syllable, *sun3*, in both (11a) and (11b) can be realized as a T3S syllable. However, the syllable belongs to different structures in the two sentences: [*zhu sun*]$_{\mathrm{N}}$ in (11a) is part of a compound noun while [*zhu sun*]$_{\mathrm{NP}}$ in (11b) is an NP that consists of a bimorphemic single noun[7] followed by an adjective. Therefore, an input of a T3S syllable does not help listeners predict the position of a phrase boundary (whether the current input is part of a compound noun or an independent NP). Hence, the occurrence of a T3S syllable may not be used to predict structural information. This is the second reason that a T3S syllable may not be utilized in predictive processing.

(11)
a. [[*Zhu2 sun3*]$_{\mathrm{N}}$ [*huo3 guo1*]$_{\mathrm{N}}$]$_{\mathrm{NP}}$

---

[7] Even though, for the most part, a monosyllabic morpheme can stand as an independent word in Mandarin Chinese, the use of bisyllabic/bimorphemic words for single referents has become firmly established in the modern language (e.g., native speakers rarely use the monosyllabic word *sun3* ‘bamboo shoot’, preferring the bisyllabic *zhu2 sun3* ‘bamboo-shoot’).

bamboo-shoot hot-pot ‘bamboo-shoot hot-pot’ / 竹筍火鍋
→[[2 2][3 1]]

b. [[*Zhu2 sun3*]$_{NP}$ [*hen3 xiang1*]$_{AP}$]$_{S}$
bamboo-shoot very fragrant ‘bamboo-shoot is fragrant’ / 竹筍很香
→[[2 2][3 1]]

3.2.3 On the optionality of T3S application

If the T3S rule were simply phonologically controlled, as mentioned in Section 3.1, the rule should apply to all T3 syllables in a sequence except the last one, as in (12).

(12) T3S as a simple phonological rule (Cheng, 1973)
Lao3 Li3 *mai3* *hao3 jiu3*.
old Li buy good wine
‘Old Li buys good wine.’ / 老李買好酒
→ 2-2-2-2-3

Empirical data, however, show many cases where the T3S rule is not applied in a T3S environment. (Henceforth, a syllable string in which an underlying T3 syllable is followed by at least one underlying T3 syllable will be termed a “T3S environment.”) As (13) shows, the sentence in (12) has multiple acceptable variations for the surface tone of the second and third syllables (including the tone sequence shown in [12] and repeated in [13d]). The acceptability of the examples from (13a) to (13d) might suggest that T3S application is random. But the unchanged T3 status of the first syllable, *lao3*, in (13e) and of the fourth syllable, *hao3*, in (13f) results in incongruity. (Henceforth, “unchanged T3” describes underlying T3 that does not change to T2 in a T3S environment.) That is, these two syllables must change to T2.[8] The variation in T3S application among the first four underlying T3 syllables in (13) shows that T3S application is not random, but must be controlled by certain factors.

(13) T3S application variations on a T3 syllable string (adapted from Cheng, 1973)
Lao3 Li3 *mai3* *hao3 jiu3*.

---

[8] In fact, *lao3* and *hao3* can maintain T3 in certain environments such as when being read aloud as isolated single words, or spoken with pragmatic emphasis or very slowly (also see Duanmu, 2007).

old Li buy good wine 'Old Li buys good wine.' / 老李買好酒
→ a. 2-3-3-2-3
→ b. 2-2-3-2-3
→ c. 2-3-2-2-3
→ d. 2-2-2-2-3
→ e. *3-2-2-2-3
→ f. *2-2-2-3-3

The problem of what factors control T3S application has been discussed for decades in the field of Chinese phonology but is still under debate (see Chen, 2000; Duanmu, 2007; Yip, 2002 for reviews). Many previous studies have tried to articulate a general linguistic rule for T3S application.

Cheng (1973) proposed that T3S application is sensitive to both syntactic structure and speech speed. He first analyzed the sentence in (13) in terms of syntactic structure as [[*Lao Li*]$_{NP}$ [[*mai*]$_{V}$ [*hao jiu*]$_{NP}$]$_{VP}$]$_{S}$, and assumed that the T3S rule is more readily applied to the left side of a word or a phrase than to its right side. Thus, the first syllable *lao* and the fourth syllable *hao* change to T2 with comparative ease. Furthermore, according to Cheng, the influence of speech speed interacts with syntactic structure, with higher speed making it easier for the T3S rule to be applied. Taking (13) as an example, when the speed is slow, T3S first applies to the smallest branches as in (14a) [[***Lao2*** *Li3*] [*mai3* [***hao2*** *jiu3*]]]. When the speed increases, T3S applies to a wider range as in (14b) [[***Lao2 Li2***] [*mai3* [***hao2*** *jiu3*]]] (taking ***Lao2 Li2*** *mai3* as a trisyllabic unit; the problem of this analysis, which is inconsistent with syntactic structure, is discussed in the next paragraph) or (14c) [[***Lao2*** *Li3*] [***mai2*** [***hao2*** *jiu3*]]] (taking ***mai2 hao2*** *jiu3* as a trisyllabic unit), according to how speakers group trisyllabic units. Finally, at a high speed, T3S applies to all possible positions in the sentence except the last one as in (14d) [[***Lao2 Li2***] [***mai2*** [***hao2*** *jiu3*]]]. Cheng's assumption can also explain the incongruity of (13e) *[[*Lao3* ***Li2***] [***mai2*** [***hao2*** *jiu3*]]] and (13f) *[[***Lao2 Li2***] [***mai2*** [*hao3* *jiu3*]]]. The unchanged T3 of *lao3* in (13e) and *hao3* in (13f) are incompatible with the priority of applying T3S to the leftmost position in a phrase.

(14) The analysis of Cheng (1973)

[[*Lao3 Li3*]$_{NP}$	[[*mai3*]$_{V}$	[*hao3 jiu3*]$_{NP}$]$_{VP}$]$_{S}$	(老李買好酒)
old Li	buy	good wine	'Old Li buys good wine.'
a. [[23]	[3	[23]]]	→Slow A (disyllabic units)
b. [[22]	[3	[23]]]	→Slow B (one more T3S, after Slow A)

c. [[23] [2 [23]]] →Medium (up to trisyllabic units)
d. [[22] [2 [23]]] →Fast (entire tree)

The problem of such syntactically based analyses (Cheng, 1973; Shen, 1994) is that they cannot provide explicit predictions for T3S application in terms of the different numbers of syllables in a phrase. For example, Cheng explained (14b), where the tone pattern of the trisyllabic group is *lao2 li2 mai3*, as “the output of application” to (14a). He did not explain why the trisyllabic unit in (14b) is allowed even though it is inconsistent with syntactic structure ([***Lao2 Li2***]$_{\text{NP}}$ [[*mai3*]$_{\text{V}}$ […]$_{\text{NP}}$]$_{\text{VP}}$), nor why the inconsistent unit in (14b) has higher priority for T3S application than the trisyllabic unit in (14c) ([…]$_{\text{NP}}$ [[***mai2***]$_{\text{V}}$ [***hao2 jio3***]$_{\text{NP}}$]$_{\text{VP}}$).

Another line of studies adopts the phonological concept of “foot” in their analyses (Chen, 2000; Dell, 2004; Duanmu, 2007; Shih, 1986, 1997), where a foot must contain at least two syllables and the head of a foot is the first (leftmost) syllable. The T3S rule applies readily to the head of a foot. (15) describes the steps of T3S application as proposed in these analyses.

(15) The foot analysis of T3S application (Duanmu, 2007)
a. Place a foot boundary before a syllable with emphatic stress.
b. Build disyllabic feet left-to-right for polysyllabic words.
c. Build feet cyclically for compounds.[9]
d. Build disyllabic feet at the lowest branches.
e. Build disyllabic feet left-to-right for other syllables.
f. Join free syllables to neighboring feet.
g. T3S applies cyclically in a foot (and optionally across feet).
h. At higher speed, T3S can apply to a larger tree node in one step.

(16) is an example of the foot analysis, again using the sentence in (12). First, there is no emphasis and no polysyllabic word in (16), so steps (15a) and (15b) are skipped. Next, feet are built for compounds or phrases, as [*Lao3 Li3*] and [*hao3 jiu3*] in (16c). Step (15d) is skipped because [*Lao3 Li3*] and [*hao3 jiu3*] are at the lowest branches of the structure and the feet have been built on them. Then, the string of T3 syllables is scanned from left to right to mark feet on the remaining syllables, as [*xiang3 mai3*] in (16e). When

---

[9] Duanmu uses “compound” to express the concept “word.” This is because, in most cases, one syllable can stand alone as a word with independent meaning in Chinese, while the words used in Duanmu’s examples are often disyllabic.

every syllable in the T3 sequence is assigned to a foot, T3S applies to each foot, as in (16g), for slow speech. In faster speech, the feet combine into bigger units for T3S application, as in (16h) and (16i).

(16) Example of a foot analysis (adapted from Chen, 2000; Duanmu, 2007; Shih, 1986, 1997)

[[Lao Li]$_{\text{NP}}$ [[xiang [mai [hao jiu]$_{\text{NP}}$]$_{\text{VP}}$]$_{\text{TP}}$]$_{\text{S}}$
老 李 想 買 好 酒
old Li want to buy good wine
‘Old Li wants to buy good wine’

	3 3	**3 3**	**3 3**	
a.	--N.A.--			→foot for emphasis
b.	--N.A.--			→foot for polysyllabic words
c.	(33)	33	(33)	→compounds
d.	--N.A.--			→foot at lowest branches
e.	(33)	(33)	(33)	→left-to-right foot building
f.	--N.A.--			→joining
g.	(23)	(23)	(23)	→slow T3S (in each foot)
h.	(23)	(22)	(23)	→medium
i.	(22)	(22)	(23)	→fast

How to generalize all the factors and make precise predictions for T3S application is still under debate. However, the existing discussions describe the general environment for the occurrence of T3S syllables: (i) the left position in a foot, word, or phrase with (ii) high speech speed. In contrast, the general environment for an unchanged T3 syllable should be: (i) the right position in a foot, word, or phrase with (ii) low speech speed.

If we compare the outputs of T3S application in different syntactic structures or in different positions in the same structure, we can see that T3S application is certainly sensitive to structure. However, do listeners utilize this sensitivity in predictive processing? From the predictive viewpoint, if every structure relates to only one specific pattern of T3S application (i.e., a one-to-one association), it should be easy for listeners to infer structural information from the T3S application pattern (i.e., if a structure X has only one T3S application pattern Y, then when part of Y is input, X can be predicted). But the problem is that, even at the same speech speed, one structure is allowed to associate with more than one pattern of T3S application in real use (e.g., 14a–d). T3S application is sensitive to structure but does not form a one-to-one association with one environment.

We have discussed T3S application in different structures. Let us now look at the case of a structure that includes only two underlying T3 syllables. If T3S application were simply triggered by the phonological environment, it would be easy to infer that a T3S syllable must be followed by an underlying T3 syllable while a surface T3 syllable must be followed by a non-T3 syllable (i.e., a complementary distribution). However, in native use, T3S application does not have a one-to-one association with its environment. A surface T3 syllable can be followed not only by a non-T3 syllable but also by an underlying T3 syllable. Therefore, the optionality of T3S application may cause interference in predicting phonological form. That is, while an input of T3S could help listeners predict an underlying T3 syllable, an input of surface T3 may not have any effect on listeners' ability to predict the tonal type of the upcoming syllable.

3.3 T3S information used in predicting the following component

Section 3.2 described the three uncertainties of T3S application, and possible situations in which the flexibility and the optionality of T3S would interfere with predictive processing. The main purpose of this dissertation is to examine whether listeners use T3S as a cue to predict the lexical structure and the tone type of the following syllable regardless of possible interference from the flexibility and the optionality of T3S application. This section introduces the experiments designed to address this question. The critical materials used in Experiment 1 and Experiment 2 employed single structures or compound structures made up of bimorphemic nouns, similar to the materials used by Hirose and Mazuka (2015, 2017). We will therefore be able to compare the results from Mandarin T3S and the Japanese CAR, and thus compare the predictive effects of prosodic information in a one-to-one association and in a non-one-to-one association.

3.3.1 The predictive effect of T3S on structure

As discussed in Section 1.5 and Section 2.4.1, the Japanese CAR has a one-to-one association with its morphosyntactic environment. The CAR obligatorily applies to every compound structure. Hirose and Mazuka (2015, 2017) found that Japanese listeners use this one-to-one association as a processing cue to predict lexical structure. That is, input with accentual change derived from the CAR (e.g., *mikan* 'tangerine' with LHH tone) helps listeners anticipate the compound structure when they are presented with candidates including a single noun structure (e.g., 'tangerine') and a compound noun structure (e.g., a compound noun modified by ‘tangerine’).

We are interested in whether Mandarin T3S has a similar predictive effect on lexical structure. As discussed in Section 3.1.3, a necessary condition for T3S application is that a T3S syllable must be followed by another underlying T3 syllable. Therefore, when Mandarin listeners hear input with tonal change derived from T3S (e.g., *zhu2 sun3→2…* ‘bamboo-shoot’), it would be reasonable for them to predict a compound structure. To test this possibility, this study’s experiments will create an environment similar to that used in Hirose and Mazuka’s (2015, 2017) experimental materials, which will include candidates named with single word structures (e.g., ‘bamboo-shoot’) and compound word structures (e.g., a compound noun modified by ‘bamboo-shoot’).

However, the flexibility of T3S application may be a potential factor that interferes with the predictive effect of T3S on lexical structure. As discussed in Section 3.2.2, it is difficult for listeners to know where to put a boundary and what type of boundary it should be after a T3S syllable. Taking (11) as an example again, a T3S syllable can occur not only before a morphemic boundary (11a) but also before a phrasal boundary (11b). That is, input including a T3S syllable such as *zhu2 sun2* ‘bamboo-shoot…’ can be the modifier of a compound noun (11a), or of a single noun (11b) that is followed by a predicate. For this reason, when a T3S syllable is the input (e.g., *zhu2 sun2* ‘bamboo-shoot…’), listeners may not anticipate a compound structure but instead retain the possibility of the single noun reading as in (11b), even when a structure such as NP + adjective is not a candidate in the specified environment.

(11)

a. [[*Zhu2 sun3*]N [*huo3 guo1*]N]NP
bamboo-shoot hot-pot ‘bamboo-shoot hot-pot’ / 竹筍火鍋
→[[2 2][3 1]]

b. [[[*Zhu2 sun3*]N]NP [*hen3 xiang1*]AP]S
bamboo-shoot very fragrant ‘bamboo-shoot is fragrant’ / 竹筍很香
→[[2 2][3 1]]

3.3.2 The predictive effect of T3S on tone type

The Japanese CAR applies to every compound regardless of the lexical tone type of the head. For this reason, listeners cannot use the CAR to predict phonological information of the head. In contrast, a T3S syllable must be followed by an underlying T3 type syllable. For this reason, the T3S rule in Mandarin Chinese offers a good testing ground for investigating predictive effects on the phonological form of upcoming material, while the CAR does not.

More concretely, in an environment that only includes choices between a compound with an underlying T3 head (e.g., *zhu2 sun3* ‘bamboo-shoot’ + *huo3 guo1* ‘hot-pot’) and a compound with a non-T3 head (e.g., *zhu2 sun3* ‘bamboo-shoot’ + *bian4 dang1* ‘lunch-box’), it would be reasonable for listeners to predict a compound with an underlying T3 head when they hear a modifier to which T3S has applied (e.g., *zhu2 sun3→2*), and to predict a compound with a non-T3 head when they hear a modifier to which T3S has not applied (e.g., *zhu2 sun3*).

However, the optionality of T3S application may interfere with the predictive effect of T3S on phonological form. As discussed in Section 3.2.3, a surface T3 syllable can be followed by a non-T3 syllable or a T3 syllable in native use. If the optionality of surface T3 diminishes the predictive effect of T3S on phonological form (and, further, diminishes its predictive effect on lexical structure, which relies on the activation of the tone type), listeners would not choose a compound with a non-T3 head when hearing input of a modifier to which T3S has not applied (in contrast to input of a T3S syllable, which may still lead listeners to predict a compound with an underlying T3 head).

This chapter has discussed the uncertainties that surround T3S application, and the possible interference these uncertainties may cause to the predictive processing of lexical structure and phonological form. Chapter 5 will present Experiment 1, which tests the predictive effects of T3S in regard to lexical structure, and Chapter 6 will present Experiment 2, which tests the predictive effects of T3S in regard to phonological form.

Chapter 4

A production task: Noun-noun compounds

4.1 Introduction

This chapter reports the results of a norming task examining how frequently the T3S syllable occurs in noun-noun compounds. This task was designed to find out whether T3S syllables frequently occur in certain lexical structures. The specific hypothesis is that Mandarin users' pronunciation frequently reflects the application of T3S in T3S environments when they are producing noun-noun compounds. The results of the task were used in the design of the materials for Experiments 1 and 2, which test the hypothesis that even though T3S application is not obligatory in native usage, listeners will use the presence of a T3S syllable as a prosodic cue to predict a following component. A noun-noun compound such as *zhu2 sun3* + *huo3 guo1* 'bamboo-shoot hot-pot' (竹筍火鍋) is composed of two bisyllabic words. The second syllable, *sun3*, and the third syllable, *huo3*, are both underlying T3 syllables, and hence the second syllable should change to a T3S syllable according to the general T3S rule. However, the second syllable is at the rightmost position of a foot boundary as well as a morphemic boundary. As discussed in Section 3.2.3, an underlying T3 syllable (in a T3S environment) in this position is most likely to retain unchanged T3. Hence, if we can confirm the preference of T3S in this position and attest to the predictability of a T3S syllable in this position on the following component, it would be reasonable to extend this hypothesis to other lexical structures.

4.2 Materials

An example of the materials used in this experiment is shown in (17). They consisted of bisyllabic single nouns, as in (17b) and (17d), and compound nouns composed of two bisyllabic single nouns (modifier + head), as in (17a) and (17c). The single nouns represent plants (e.g., *zhu2 sun3* 'bamboo-shoot' [竹筍]) while the compound nouns represent animals modified by plants (*zhu2 sun3* + *ru3 niu1* 'bamboo-shoot cow' [竹筍乳牛]). The tone type of the second syllable was manipulated to be T3, as in (17a) and (17b), or non-T3, as in (17c) and (17d). The tone type of the third syllable in the compounds, as in (17a) and (17c), was always T3.[10] Note that the tonal notation, such as

---

[10] The single/modifier/head nouns used in the experiments of this study are all bisyllabic nouns. The critical syllable positions for manipulating tone type (T3 or non-T3) are: the second syllable position in single/modifier nouns (e.g., *sun3* in *zhu2 <u>sun3</u>* 'bamboo-shoot') and the first syllable position in

T3 + T3 in the T3 compound in (17a), only represents the underlying tones, and does not represent the actual pronunciation. The actual pronunciation of the modifier (whether it is pronounced as T3S or unchanged T3) is the object of observation in this task. The only results to be analyzed are those for the T3 compounds as in (17a) because this is the only environment where T3S may occur. Words of the types in (17b–d) were used as fillers. A list of the experimental items is attached in **Appendix A**.

(17)

a. T3 compound (critical pattern)	
zhu2 sun3 + ru3 niu2	‘bamboo-shoot cow’ / 竹筍乳牛
b. T3 single (filler)	
zhu2 sun3	‘bamboo-shoot’ / 竹筍
c. non-T3 compound (filler)	
xiang1 jiao1+ ru3 niu2	‘banana cow’ / 香蕉乳牛
d. non-T3 single (filler)	
xiang1 jiao1	‘banana’ / 香蕉

4.3 Procedures

A paper list written in traditional Chinese characters was prepared. It included twenty-four critical items with T3 compounds and thirty-six filler items (six items of T3 single nouns + twenty-four items of non-T3 compounds + six items of non-T3 single nouns).

Participants were given instructions before they completed the task, which was presented on a computer. First, they were told that the whole experiment was a monster-hunting game.[11] The goal was to get as many imaginary monsters as they could at different stages (such as by pronouncing the monsters’ names or using the computer mouse to click on an image on the screen). They were instructed to read the paper list, which included all the single nouns and some combinations of compound nouns (represented by both visual objects and Chinese characters) before starting the experiment. This procedure ensured the plausibility of the novel compounds of plant and animal names, reducing the effect of the familiarity of existing words.

---

head nouns (e.g., *ru3* in *ru3 niu2* ‘cow’). Henceforth, unless otherwise specified, “single noun/modifier noun/head noun” represent bisyllabic single/modifier/head nouns with the manipulation on the specified syllable position.

[11] This instruction applies to all the experiments reported in this dissertation, and the same explanation was given to the participants of each experiment.

Because speech speed is considered a factor in T3S application, participants were asked to read the paper list twice at different speeds: first, at a natural speed and then at a slow speed with intentional pauses between each syllable. This procedure produced the data for two speed conditions, which were then analyzed in order to examine how speech speed affects T3S application in production. Their pronunciations were recorded by Praat at 44,100 Hz. The annotation of tone type was done by the author.

4.4 Participants

Thirteen native Mandarin Chinese speakers from Taiwan who live in the Tokyo area, none of whom had any visual or hearing impairment, participated in the experiment for payment.

4.5 Results

The results were analyzed in a generalized linear mixed model with a backward selection procedure using the lme4 package in R. The natural speed condition was coded as 1, and the slow speed with intentional pause condition was coded as 0. The logit of (T3S syllable/unchanged T3 syllable) was calculated as a dependent variable; speed was treated as an independent fixed factor, and subject and item were treated as random factors.

The proportions of T3S syllables in the two speed conditions are shown in **Figure 4**. As the figure shows, 93.9% of the T3 compounds at natural speed and 87.4% of them at slow speed were pronounced as T3S syllables. The intercept of this model is significantly different from 0 ($\beta = 2.7832$, $SE = 0.2860$, $z = 9.733$, $p < 0.00$ ***), which means that the proportion of T3S syllable/unchanged T3 syllable differs from the chance level. Namely, the production of T3S syllables was significantly higher than that of unchanged T3 syllables. This result suggests that T3S application affects pronunciation in noun-noun compound structures with high frequency in native usage.

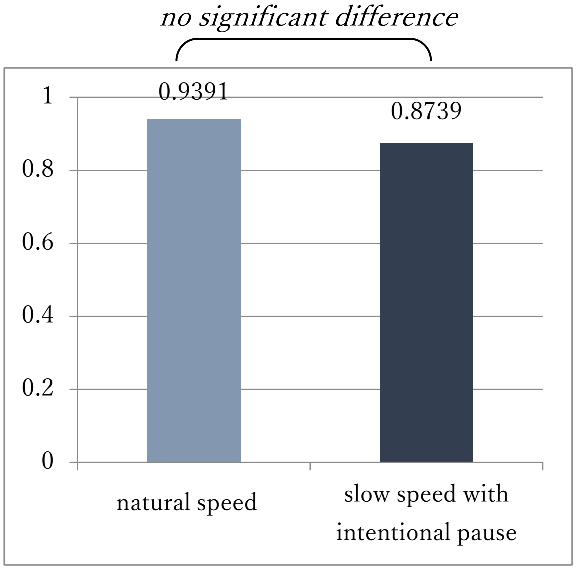


*Figure 4***.** The proportion of T3S syllables to T3 syllables in compounds produced in the pronunciation task.

Furthermore, there was no main effect of speed (β = 0.02584, *SE* = 0.6547, *z* = 0.395, *p* = 0.693), showing that there was no difference between natural speed and slow speed. This suggests that the application of the T3S rule is steady and does not decrease even with slow speed. Therefore, the effect of speed can be excluded from consideration in the following experiments.

4.6 Discussion

We hypothesized that if a T3S syllable is prominent in a certain lexical structure, listeners will use the T3S syllable as a cue to predict the following structure, regardless of the ambiguities that may result from the nonobligatory and flexible nature of T3S application. The results of the pronunciation task reported in this chapter confirm the hypothesis: Mandarin users' pronunciation frequently reflects the application of T3S in T3S environments when they are producing noun-noun compounds. Next, Chapter 5 examines whether the presence of a T3S syllable helps listeners to anticipate the following structure.

Chapter 5

Experiment 1: The anticipatory effect of T3S on lexical structure

5.1 Introduction

This chapter's aim is to examine whether T3S syllables are used as a prosodic cue to predict lexical structure, even though T3S application does not have a strict one-to-one association with specific lexical structures.

As discussed in Section 3.1, there are two necessary conditions of T3S application, restated here in (18a) and (18b).

(18)
a. Another syllable must follow the T3S syllable.
b. The syllable following the T3S syllable must be of the T3 type.

This chapter focuses on the predictive effect related to (18a) while Chapter 6 examines the predictive effect related to (18b).

If the two conditions in (18) were the only principles affecting processing in a T3S environment, it would be easy for listeners to associate input of a T3S syllable with a compound word while associating input of a surface T3 syllable with a single word, when the choice for the listener is between a compound word where a T3S syllable is at the end of the modifier and a lexical T3 syllable is at the start of the head (e.g., *zhu2 sun3* 'bamboo-shoot' [竹筍] + *ru3 niu2* 'cow' [乳牛] →*zhu2 sun2 ru3 niu2*), and a single word where a lexical T3 syllable is at the end (e.g., *zhu2 sun3* 'bamboo-shoot' [竹筍]). This situation could further result in the inference that the occurrence of a T3S syllable or a surface T3 syllable would trigger predictions on lexical structure.

However, the uncertainties of T3S application in native uses challenge such an inference. The flexibility of T3S application is one source of uncertainty. As discussed in Section 3.2.2, unlike the Japanese CAR, which only applies to compound structures, the T3S rule can apply to various structures. For example, a T3S syllable can be associated with either a compound word (e.g., *zhu2 sun3* 'bamboo-shoot' + *ru3 niu2* 'cow' → [[*zhu2* ***sun2***]$_{\mathrm{N}}$[*ru3 niu2*]$_{\mathrm{N}}$]$_{\mathrm{NP}}$) or a structure including a single word followed by a predicate (e.g., *zhu2 sun3* 'bamboo-shoot' + *hao3 chi1* 'very delicious' → [[*zhu2* ***sun2***]$_{\mathrm{N}}$ [*hao3 chi2*]$_{\mathrm{AP}}$]$_{\mathrm{S}}$). This flexibility leads to ambiguity; hence, the possibility of a single word with a T3S syllable may be triggered even if the single word is not a candidate in the context (thus contrasting with a context where the choice is between a single word with a lexical T3 syllable and a compound word with a T3S syllable). If this is true, input of a T3S syllable may not create an anticipatory effect regarding an upcoming compound word.

The optionality of T3S application is another source of uncertainty that may interfere with its predictive effect on lexical structure. As discussed in Section 3.2.1, the T3S rule is allowed NOT to apply in native uses. For example, *zhu2 sun3* 'bamboo-shoot' [竹筍] + *ru3 niu2* 'cow' [乳牛] has two realizations: one with a T3S syllable as *zhu2* ***<u>sun2</u>*** *ru3 niu2* and the other with a lexical T3 syllable as *zhu2* ***sun3*** *ru3 niu2*. In other words, a surface T3 syllable could be ambiguous as to whether it is part of a single structure (e.g., a lexical T3 syllable in *zhu2* ***sun3*** 'bamboo-shoot') or a compound structure (e.g., an unchanged T3 syllable in *zhu2* ***sun3*** *ru3 niu2* 'bamboo-shoot cow') before the listener has encountered a head noun (e.g., *ru3 niu2* 'cow'). Therefore, input of a surface T3 syllable may not be a sufficient cue to exclude the compound interpretation.

If the uncertainties of T3S application override the predictive effect of the tonal information, the predictive effect may be cancelled out in lexical processing. However, as confirmed in the experiment reported in Chapter 4, native speakers show a strong preference for producing T3S syllables in noun-noun compounds. It is conceivable that this preference for T3S syllables in compounds means that T3S can be used as a cue to predict lexical structure despite the uncertainties of T3S application. That is, even though T3S application does not have a strict one-to-one association between the tonal type and the lexical structure as the CAR does, listeners may still positively use the occurrence of a T3S syllable or a surface T3 syllable as a cue to predict a compound or a single structure, respectively. This chapter reports the results of the visual world paradigm experiment conducted to examine this hypothesis.

5.2 Materials

The materials of this experiment include auditory stimuli and corresponding visual displays in four experimental conditions:

a. T3/single condition
b. T3/compound condition (T3S is applied)
c. non-T3/single condition
d. non-T3/compound condition

5.2.1 Auditory stimuli

A sample set of the auditory stimuli is shown in (19). (See **Appendix B** for the full list of items used in Experiment 1.) The conditions manipulate structure and tone type. Two conditions employ single nouns that name individual plants, as in (19a) and (19c). Two conditions employ compound nouns consisting of a modifier noun and a head noun, as in (19b) and (19d). The compounds represent animals modified by plants.

The underlying tone type of the modifiers (and single nouns) is manipulated as the T3 type, in conditions (a) and (b), or as the non-T3 type, in conditions (c) and (d). The heads of the compounds, as in (b) and (d), both start with a T3 type syllable. Accordingly, the T3/compound condition (b) is the only condition that undergoes T3S application (i.e., it has two underlying T3 syllables in sequence).

(19)

a. T3/single

[*zhu2* 竹 *sun3* 筍] ‘bamboo-shoot’

b. T3/compound (T3S is applied)

[*zhu2* 竹 *sun3* 筍]+ [*ru3* 乳 *niu2* 牛]→ 2-2-3-2[12] ‘bamboo-shoot cow’

c. non-T3/single

[*xiang1* 香 *jiao1* 蕉] ‘banana’

d. non-T3/compound

[*xiang1* 香 *jiao1* 蕉] [*ru3* 乳 *niu2* 牛] ‘banana cow’

---

[12] “[*Zhu2* 竹 *sun3* 筍] + [*ru3* 乳 *niu2* 牛]” represents the underlying tone type; “2-2-3-2” represents the actual pronunciation in the prerecorded auditory stimuli.

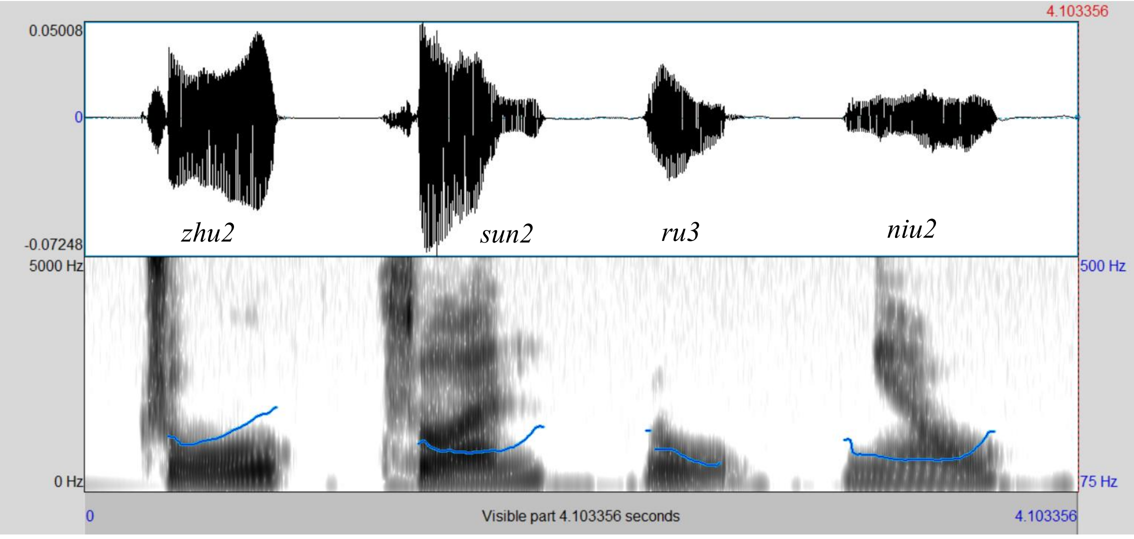


Figure 5. Spectrogram of an example of the auditory stimuli used in Experiment 1 (*zhu2 sun2 ru3 niu2* 'bamboo-shoot cow').

The auditory stimuli were recorded by the author, who is a female native speaker of Mandarin Chinese from Taiwan. The stimuli were pronounced at a speed of one syllable/1000 ms with pauses (using a stopwatch to align the onset of each syllable at 1000 ms intervals as shown in **Figure 5**). This manner of recording allowed for the control of vowel duration[13] and for the insertion of artificial blanks.[14]

5.2.2 Visual display

Each item's visual display was divided into eight areas, and each area was assigned a visual object. **Figure 6** shows an example of a display used for the auditory stimuli of the T3 conditions (a) and (b). The middle bottom object {bamboo-shoot}[15] (竹筍 *zhu2*

---

[13] Listeners tend to predict a longer word when hearing a shorter vowel, and the acoustic difference is sufficient to permit the detection of word boundaries in comprehension (see Davis, 2000 for a review). In addition, in the recordings of the auditory stimuli made for this experiment's pilot tests, the vowel duration was found to differ depending on the presence or absence of a following syllable in fast speech without pauses (for example, *jiao1* was longer in *xian1 jiao1* than in *xian1 jiao1 ru3 niu2*). For these reasons, pauses were inserted after every syllable to control the vowel duration.

[14] The onset of each syllable was aligned using a stopwatch because artificial blanks manipulated by recording applications such as Praat may delete some acoustic information and mislead listeners into processing every syllable as an individual segment (syllable perception) rather than as part of a word (lexical processing).

[15] The plant or animal names in curly brackets such as {bamboo-shoot} represent the visual object shown in the visual display. The tonal description after the curly brackets (e.g., 竹筍 *zhu2 sun3*) represents the underlying tone rather than the actual pronunciation.

sun3) is the [Target Single Object], which was the correct answer for the auditory T3/single condition (a). The second-row rightmost object {bamboo-shoot cow} (竹筍乳牛 *zhu2 sun3 + ru3 niu2*) is the [Target Compound Object], which was the correct answer for the auditory T3/compound condition (b).

Each visual display also contained a [Competitor Compound Object]; for example, in **Figure 6**, the leftmost bottom object {bamboo-shoot raccoon} (竹筍浣熊 *zhu2 sun3 + wan3**[16] *xiong2*) is a competitor target in the auditory T3/compound condition (b). The presence of such a [Competitor Compound Object] should show whether listeners are able to anticipate the compound structure at an early stage; that is, before encountering the segmental information of the head noun. This is because both {bamboo-shoot cow} (*zhu2 sun3 + ru3 niu2*) and {bamboo-shoot raccoon} (*zhu2 sun3 + wan3* xiong2*) include two underlying T3 syllables in sequence at the same position. If the T3S syllable in the auditory stimulus *zhu2 sun2…* is a sufficient cue to predict compound structure, we should observe increases in fixations on both the [Target Compound Object] and the [Competitor Compound Object] before the participants hear the head noun *ru3 niu2* of the auditory stimuli.

To balance the number of single and compound nouns represented in the visual displays, each display included six more visual objects corresponding to nouns (three singles and three compounds) with T3 syllables. For example, the visual display in **Figure 6** includes: {sweet-potato} (番薯 *fan1 shu3*), {sweet-potato cow} (番薯乳牛 *fan1 shu3 + ru3 niu2*), {sweet-potato raccoon} (番薯浣熊 *fan1 shu3 + wan3* xiong2*), {cow} (乳牛 *ru3 niu2*), and {raccoon} (浣熊 *wan3* xiong2*).

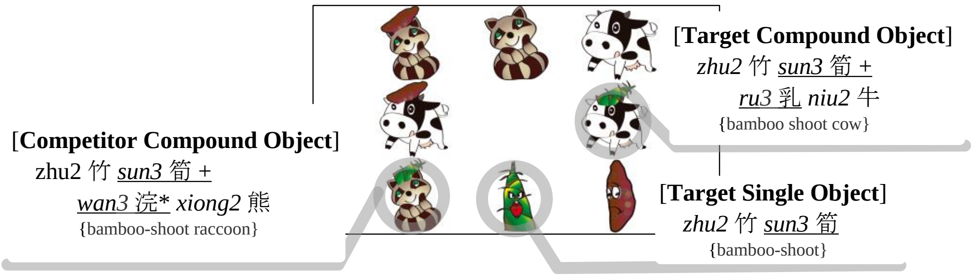


Figure 6. Example of a visual display for the auditory T3 conditions, with 'bamboo-shoot'

---

[16] A syllable marked by an asterisk (*) has a different pronunciation in Taiwan Mandarin and Beijing Mandarin. These are the three items used in this study's experiments that are different in the two dialects, with the Taiwan pronunciation shown to the left and the Beijing pronunciation shown to the right of the slashes: (i) 浣熊 'raccoon': *wan3 xiong2* / *huan4 xiong2*; (ii) 企鵝 'penguin': *qi4 e2* / *qi3 e2*; (iii) 蘿蔔 'radish': *luo2 bo1* / *luo2 bo0*.

as the first constituent, in Experiment 1.

When visual objects corresponding to the T3 type are used as the visual targets, visual objects corresponding to the non-T3 type such as {banana} (香蕉 *xiang1 jiao1*), {banana cow} (香蕉乳牛 *xiang1 jiao1+ ru3 niu2*), and {banana raccoon} (香蕉浣熊 *xiang1 jiao1+ wan3* xiong2*) are not appropriate as distractors in the same visual display, because it would be possible for listeners to eliminate options using the tone type information (T3 or non-T3) of the first constituent. For this reason, visual target objects of the non-T3 type (e.g., {banana} *xiang1 jiao1*) were presented with visual distractor objects of the non-T3 type (e.g., {green pepper} *qing1 jiao1*), as shown in **Figure 7**.

Figure 7. Example of a visual display for the auditory non-T3 conditions, with ‘banana’ as the first constituent, in Experiment 1.

5.2.3 Filler items

The filler auditory stimuli were of non-T3 types and were single word names for animals such as *hu2 die2* ‘butterfly’ [蝴蝶] and *da4 xiang4* ‘elephant’ [大象]. The visual displays for the filler auditory items were similar to the visual displays used for the critical auditory items. The only differences were that the [Target Compound Object] and [Competitor Compound Object] were replaced by two non-T3 single nouns for animals, one of which was the target while the other was a distractor. For example, the visual display in **Figure 6** was changed to be used for the filler auditory item *hu2 die2* ‘butterfly’ by replacing {bamboo-shoot cow} with {butterfly} and {bamboo-shoot raccoon} with {elephant}.

The materials included 24 critical auditory items and 24 filler auditory items as well as their corresponding visual displays. The critical auditory items were divided into four lists in a Latin-square design. The same filler auditory items were used in each list. Every list included 48 trials. Each participant was assigned to one of the four lists by Latin squares of order.

5.3 Participants

Twenty-four native Mandarin Chinese speakers from Taiwan who lived in the Tokyo area, and who had no visual or hearing impairment, participated in this experiment for payment (18 females, 6 males; mean age 25.25 ± 4.87; age range 18–38). The mean of the period in which the participants had lived outside of Taiwan was 1.83 ± 1.89 years (range: 0.1–8). All of them had finished 12 years of compulsory education in Taiwan.

5.4 Procedures

Before they began the main VWP experimental task, participants received pre-experimental instructions (the same as those for the production task reported in Chapter 4). The participants were told that all the objects they would hear about and see were imaginary monsters. This instruction was intended to ensure the plausibility of combining the plant and animal names in compound nouns. The participants were further told that even though all the objects were monsters, the word *guai4 shou4* 'monster' [怪獸] would be omitted in all auditory stimuli. This was done to prevent the participants from wrongly analyzing an intended single condition (e.g., $[\text{bamboo-shoot}]_{\text{N}}$) as the modifier of a compound (e.g., $[[\text{bamboo-shoot}]_{\text{N}}\ [\text{monster}]_{\text{N}}]_{\text{NP}}$). The participants were also told that every monster had eyes, so that they would distinguish the modifier reading in compound objects (e.g., the {bamboo-shoot} part of a [Target Compound Object] such as a {bamboo-shoot cow} has no eyes) from the single noun reading (e.g., the {bamboo-shoot} in the [Target Single Object] has eyes). Next, they were asked to read a written list of the names of all of the objects used as experimental items. This procedure made them pronounce all of the words they would pronounce during the experiment, in order to prevent ambiguity arising from surface rising tone, which can be associated with both T2 and T3 morphemes (as discussed in Section 3.4.1). More specifically, this procedure was intended to prevent the participants from wrongly associating the experimentally intended T3S sound with a T2 meaning.

The main VWP experiment used E-Prime 2.0 + Tobii extension to present stimuli and Tobii TX300 to record participants’ eye movements. The procedure of one trial is illustrated in **Figure 8**. Participants sat in front of a Tobii screen of 23 inches with 1928 × 1080 pixels. After 1000 ms of fixation, the visual display was first presented for 2500 ms to ensure enough time for participants to retrieve prosodic and lexical information of the visual objects on display. Then, the auditory stimulus was presented by the sound speakers on two sides of the Tobii TX300 screen. The participants were asked to click on the correct answer using the computer mouse as quickly as possible (i.e., they were told it was not necessary to wait until the auditory stimuli finished). The participants received feedback at the end of each trial, telling them whether their answer was correct. Participants’ eye movements were recorded from the onset of the auditory stimuli until the mouse click.

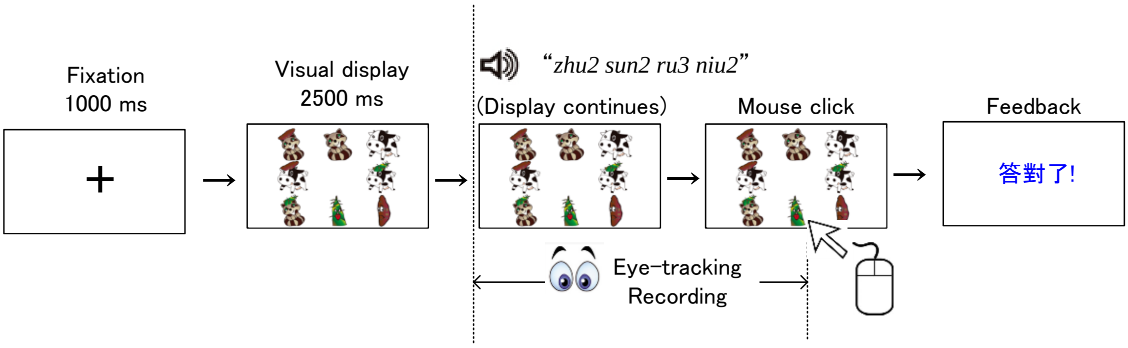


Figure 8. Illustration of the procedure of Experiment 1.

5.5 Predictions

If the hypothesis that Mandarin listeners use the occurrence of a T3S or a surface T3 syllable as a cue to predict lexical structure regardless of the uncertainties of T3S application is correct, then this experiment should show predictive effects in three different ways. First, input of a T3S syllable should lead participants to anticipate a compound interpretation; therefore, there should be more looks to compound structure objects when the participants hear input with a T3S syllable. More specifically, compared to the baseline T3/single condition (a), the T3/compound condition (b) should lead to an increase in looks to both the [Target Compound Object] and the [Competitor Compound Object]. Furthermore, if the predictive effect relies on the input of a T3S syllable in the modifier but not on the segmental information of the head, the increase of looks to the [Target Compound Object] and the [Competitor Compound Object] in the T3/compound condition (b) should start earlier than the auditory onset of the head noun (2000 ms).

Second, Hirose and Mazuka (2015) demonstrated that when participants have anticipated compound structure when hearing the accentual change derived from the Japanese CAR, the compound representation the listeners have already built facilitates the next process of identifying the head noun and computing the meaning of the compound. Therefore, their participants distinguished target compound objects from competitor compound objects more rapidly in conditions with a prosodic cue compared to conditions with no prosodic cue. If T3S has the same effect, leading to earlier identification of the head noun, the time point for the disambiguation of the "final correct answer" from the "other candidate" should be earlier in the T3S compound condition (b) than in the non-T3 compound condition (d). That is, the time point of the divergence between the [Target Compound Object] and the [Competitor Compound Object] when hearing a T3/compound should be earlier than when hearing a non-T3/compound.

Third, it is conceivable that not only a T3S syllable but also a surface T3 syllable is informative for predicting lexical structure. That is, input of a surface T3 syllable might lead listeners to anticipate a single noun interpretation. If so, there should be more looks to the [Target Single Object] in the T3/single condition (a) than in the T3/compound condition (b).

5.6 Data Analysis

The sampling rate for the eye movement tracking was 300 Hz (one sampling point/3–5 ms). The eye movement types that the Tobii X300 tracker records are basically fixations and saccades. The saccades were excluded (because the X-Y coordinates could not be recognized) as were the data points that the tracker could not classify (e.g., when the tracker lost track of the participants' pupils). Next, the X-Y coordinates of fixations were mapped onto the eight areas in the visual display. The fixation times on the given areas of interest (AOI) were summed for the given time intervals in each analysis.

This section reports the results from two methods of statistical analysis. One is permutation cluster analysis (also referred to as randomization tests or nonparametric statistical testing; Maris, 2012; Maris & Oosternveld, 2007), for observing the explicit divergence time interval between any two conditions of interest. Permutation cluster analysis, however, is unsuited for cases with more than two conditions and for examining the interactions of multiple factors. Therefore, the second method employed in this study is linear mixed effect regression modeling (LME; Baayen, Davidson, & Bates, 2008), for evaluating the interaction effect between the experiments' multiple fixed factors and estimating the amount of variation between items and participants. The results of the LME analysis will be reported for each syllabic interval (every 1000 ms).[17]

5.6.1 Permutation analysis

Each trial lasted approximately 4000 ms. The data were broken down into 20 ms time windows, and the mean of fixation times by subject to the given AOI object (or combination of them) was calculated for each 20 ms window. The *p*-value statistics on the difference between the two conditions of interest were calculated by using a simple linear model fit. Any adjacent windows where the *p*-statistic reached significance at the 0.05 level in the same direction were clustered together, and the sum of the *t*-values for every individual 20 ms window was calculated within these clustered windows (longer clusters usually have stronger *t*-values) as new statistics that could be compared to the permutated distribution results from later simulations. Next, the labels of the two conditions on the original data points in significant clusters (*p*-values lower than 0.05) were randomly exchanged 1,000 times for every 20 ms window to create a permutation distribution of *t*-values (only extracting the maximum *t*-value of each simulation to lower the false alarm rate). If the summed *t*-value of the observed cluster in the prior step was greater than 97.5% (= 0.05 level of significance in a two-tailed test) of the permutation distribution obtained from the resampling simulation, the significance of the difference in the observed cluster was determined.

5.6.2 LME analysis

The dependent variable used in the LME (which was modeled using the lmer4 package in the R program) was log odds, which express a fixation likelihood of looking against not-looking at the given AOI(s) (i.e., “empirical logit”; Gart, Pettigrew & Thomas, 1985): the logarithm of [AOI + 0.5/ALL - AOI + 0.5] (±0.5 is an adjustment for preventing a zero numerator in logarithms). The models included tone (T3 = 1; non-T3 = 0) and structure (single = 1; compound = 0) as predictor factors. The coding was

---

[17] Permutation analysis and LME analysis use two different kinds of dependent variables. In this study, using means of fixation time by subject as a dependent variable, the permutation analysis provided a more intuitive way to represent a multiplicative increase from looks to the [Target Compound Object] alone (**Figure 10**) to looks to the [Target Compound Object] + [Competitor Compound Object] combined (**Figure 9**). In contrast, the LME analysis, using empirical logit, can compare the probabilities of looking and not looking at an object for the dependent variable, which was useful for lessening the skewness of the data distribution resulting from data loss (saccades or unclassified data points).

normalized. The random slope structure included subject and item. Model selection used a backward stepwise process (Bates, Kliegl, Vasishth & Baayen, 2015), and the *p*-values were obtained by likelihood ratio tests of the full model against the model without the effect in question (lmerTest package; Kuznetsova, Brockhoff & Christensen, 2017).

5.7 Results

This section reports the results of three sets of analyses. The first analysis investigates the fixations on [Target Compound Object] + [Competitor Compound Object] combined to examine the anticipatory effect of a T3S syllable on compound structure and observe whether the effect was revealed earlier than the input of head noun. The next analysis looks at the difference between fixations on [Target Compound Object] and fixations on [Competitor Compound Object] to confirm when the designated compound target is identified. The third analysis considers the fixations on [Target Single Object] to see how the information of a surface T3 syllable is utilized to anticipate a single noun interpretation.

5.7.1 Looks to [Target Compound Object] + [Competitor Compound Object]

Figure 9 shows the proportion of fixations on [Target Compound Object] and [Competitor Compound Object] combined in the four auditory conditions. To intuitively represent the divergence time point, the proportions of fixations in the four auditory conditions are split into two panels: the upper panel presents the two T3 conditions (a) and (b), while the bottom panel presents the two non-T3 conditions, (c) and (d). The colors of the lines present the structure type: the solid red lines for compounds, and the dashed blue lines for single nouns. Taking an item for a concrete example, the red line in the upper panel represents the sum of the proportion of fixations on {bamboo-shoot cow} and the proportion of fixations on {bamboo-shoot raccoon} when hearing *zhu2 sun2 ru3 niu3* ‘bamboo-shoot cow’ (T3/compound condition), while the blue line in the upper panel shows the sum of proportions of fixations on these two objects when hearing *zhu2 sun3* ‘bamboo-shoot’ (T3/single condition). The gray shading indicates the time spans showing significant differences, identified by the permutation cluster analysis in the comparisons between single and compound auditory conditions.

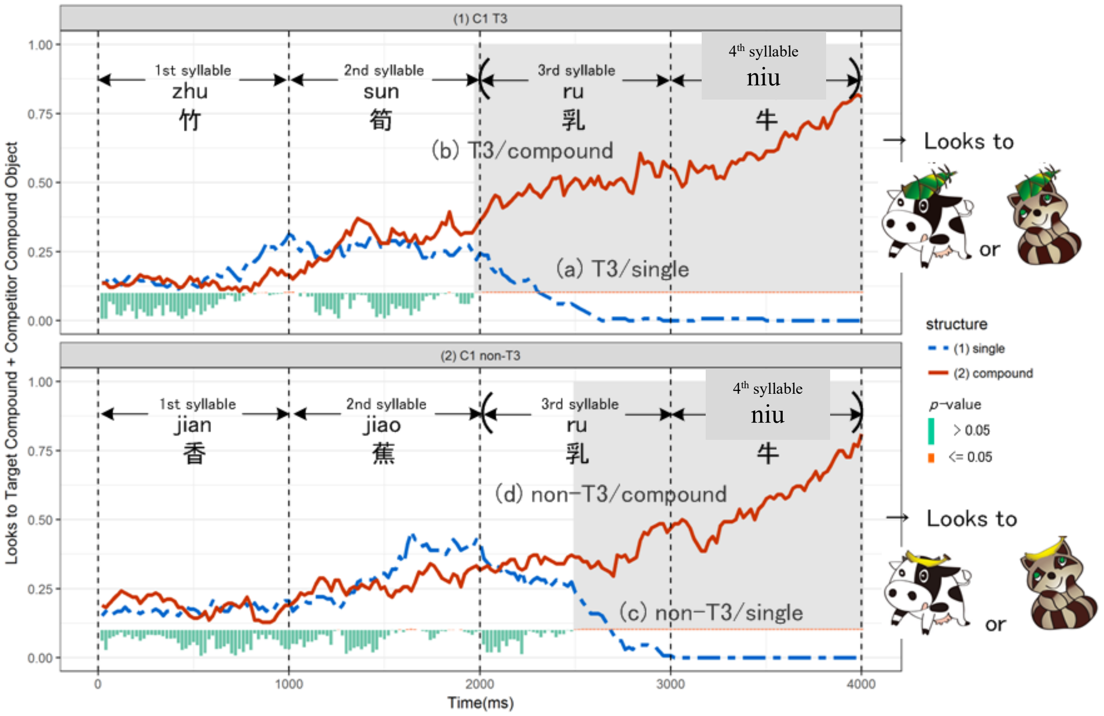


Figure 9. Proportion of fixations on [Target Compound Object] + [Competitor Compound Object].

5.7.1.1 Permutation analysis

The results of the permutation analysis show that in the unambiguous T3 condition set there were significantly more looks to compound objects in the T3/compound condition (b), that is, when T3S was applied, compared to the T3/single condition (a), which kept surface T3, in the 1980–4000 ms window (cluster *t*-value = 1232.372, cluster *p*-value < 0.001). In the ambiguous non-T3 condition set, there were significantly more looks in the non-T3/compound condition (d) compared to the non-T3/single condition (c) in the 2500–4000 ms window (cluster *t*-value = 850.951, cluster *p*-value < 0.001). The divergence time point in the unambiguous T3 condition set was earlier than that in the ambiguous non-T3 condition set (1980 ms < 2500 ms). Furthermore, the divergence of T3 conditions appeared earlier than the onset of the head noun (1980 ms < 2000 ms).

5.7.1.2 LME analysis

The LME analysis determined the interaction effect within each syllable interval (**Table 1**). The length of each syllable (including a natural pause) is 1000 ms. No effect was observed in the 2nd syllable. The 3rd syllable showed an interaction effect between tone and structure ($\beta = -3.028$, $t = -4.756$, $p < 0.01$***), reflecting that the difference within the unambiguous T3 condition set ($\beta = -5.810$, $t = -9.297$, $p < 0.01$***) was larger than that within the ambiguous non-T3 condition set ($\beta = -2.764$, $t = -4.717$, $p < 0.01$***). This result indicates that participants' fixations on candidate compound objects increased most quickly in the T3/compound condition (b) while the fixations decreased most quickly in the T3/single condition (a). While the stimuli of the non-T3/compound condition (d) also led to a higher number of looks to candidate compounds than the non-T3/single condition (c), the degree of difference is relatively smaller than that in the unambiguous T3 condition set.

Table 1

Summary of the Mixed Logit Model of Fixations on [Target Compound Object] + Target Single Object]

		β	error	*t*	*p*
2nd syllable (1000–2000 ms)	(Intercept)	-1.320	0.323	-4.092	
	Tone	0.164	0.332	0.495	0.62
	Structure	-0.005	0.287	0.017	0.99
	Tone×Structure	0.787	0.574	-1.371	0.17
3rd syllable (2000–3000 ms)	(Intercept)	-1.567	0.261	-6.002	
	Tone	0.106	0.377	0.282	0.77
	Structure	-4.291	0.480	-8.931	<0.01***
	Tone×Structure	-3.028	0.637	-4.756	<0.01***
(a) vs. (b)	Structure	-5.810	0.625	-9.297	<0.01***
(c) vs. (d)	Structure	-2.764	0.586	-4.717	<0.01***
4th syllable (3000–4000 ms)	(Intercept)	1.010	0.278	3.628	
	Tone	1.443	0.333	4.329	<0.01***
	Structure	-5.308	0.367	-14.470	<0.01***
	Tone×Structure	1.371	0.609	2.250	0.03*
(a) vs. (b)	Structure	-4.633	0.525	-8.816	<0.01***
(c) vs. (d)	Structure	-5.985	0.375	-15.976	<0.01***

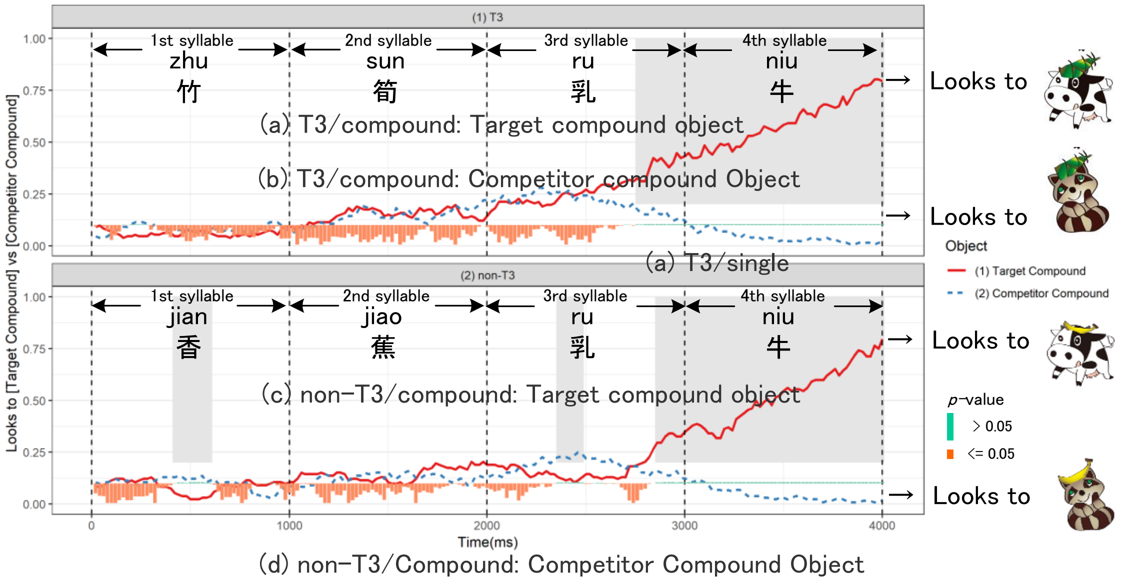
. $p=0.1$, *$p = 0.05$, **$p = 0.01$, ***$p = 0.001$

5.7.2 Looks to [Target Compound Object] versus [Competitor Compound Object]

This section examines whether the presence of a T3S syllable influences the anticipation of a compound by analyzing differences between fixations on the [Target

Compound Object] and on the [Competitor Compound Object] when participants heard the corresponding auditory stimuli: the T3/compound condition (b) versus the non-T3/compound condition (d). If the T3S rule has the same effect on predictive processing of the head noun as the CAR has, the time point of distinguishing the [Target Compound Object] (i.e., the correct answer) from the [Competitor Compound Object] (i.e., the incorrect answer) should be earlier in (b), the T3/compound condition (where T3S is applied) than in (d), the non-T3/compound condition.

Figure 10 shows the proportion of fixations on the [Target Compound Object] versus the [Competitor Compound Object] in the T3/compound and non-T3/compound auditory conditions.[18] The solid red lines represent the fixations on the [Target Compound Object] whereas the dashed blue lines represent those on the [Competitor Compound Object]. The upper panel shows the cases when participants heard auditory T3/compound stimuli (note that the T3/single auditory condition is not included) while the bottom panel presents the cases when they heard auditory non-T3/compound stimuli. Taking the visual objects shown in **Figure 10** for examples, {bamboo-shoot cow} is the correct answer while {bamboo-shoot raccoon} is the incorrect answer to the auditory T3/compound stimulus *zhu2 sun2 ru3 niu2*. A divergence of these two lines is expected after encountering the head noun (i.e., from the 3rd syllable).



---

[18] Note that the conditions discussed in Section 5.7.1 and Section 5.7.3 are the four types of auditory stimuli (the dependency was the proportion of fixations on the same type(s) of visual object). However, the conditions discussed in Section 5.7.2 are combinations of two types of visual object and two types of auditory stimulus.

Figure 10. Proportion of fixations on [Target Compound Object] vs. [Competitor Compound Object].

5.7.2.1 Permutation analysis

First, the permutation analysis showed that in the T3/compound conditions, there were significantly more looks to the [Target Compound Object] than to the [Competitor Compound Object] in the 2760–4000 ms window (cluster *t*-value = 694.217, cluster *p*-value < 0.001). In the non-T3/compound conditions, the first two significant intervals were unexpected, where more fixations were observed on the [Competitor Compound Object] than the [Target Compound Object] in the 420–600 ms window (cluster *t*-value = 28.053, cluster *p*-value < 0.001) and in the 2360–2480 ms window (cluster *t*-value = 19.807, cluster *p*-value = 0.008). The possible effect of these two windows will be discussed in Section 5.8.2, along with the findings reported in the next section. The third significant interval was consistent with the prediction: there were more looks to the [Target Compound Object] than to the [Competitor Compound Object] in the 2860–4000 ms window (cluster *t*-value = 599.599, cluster *p*-value < 0.001). The divergence point in the T3/compound conditions was 100 ms earlier than that in the non-T3/compound conditions.

5.7.2.2 LME analysis

The results for every syllable interval (1000 ms) in the LME logit model are reported in **Table 2**. The dependent variable was the empirical logarithm of looking against not-looking to the AOI ([Target Compound Object] or [Competitor Compound Object]). Models included tone (T3 = 1; non-T3 = 0) and object ([Target Compound Object] = 1; [Competitor Compound Object] = 0) as predictor factors. The random slope structure included subject and item. A marginal main effect of tone was found in the 2nd syllable. In the 3rd syllable, no interaction effect but the main effect of tone was revealed. In the 4th syllable, no interaction effect but the main effect of object was revealed. Further modeling for the 2760–4000 ms window (2760 ms is the divergence point in the T3/compound conditions) tested whether the absence of an interaction effect at the 3rd syllable was due to the unexpected effect in the 2360–2480 ms window. It showed the same tendency as seen for the 4th syllable, where no interaction was found, but there was a main effect of object ($\beta = 6.959$, $t = 14.560$, $p = < 0.01$***), indicating that there was an advantage of [Target Compound Object] over [Competitor Compound Object] regardless of the tone type of the auditory stimuli.

Table 2

Summary of the Mixed Logit Model of Fixations on [Target Compound Object] and on [Competitor Compound Object] for Every Syllable Interval

		β	error	*t*	*p*
1st syllable	(Intercept)	-4.972	0.162	-30.64	
(0–1000 ms)	Tone	-0.173	0.276	-0.624	0.533
	Object	-0.305	0.276	-1.104	0.270
	Tone×Object	-0.073	0.553	-0.131	0.896
(a) vs. (b)	Object	-0.342	0.384	-0.889	0.375
(c) vs. (d)	Object	-0.269	0.397	-0.677	0.499
2nd syllable	(Intercept)	-3.633	0.259	-14.043	
(1000–2000 ms)	Tone	0.618	0.311	1.986	0.053 .
	Object	0.113	0.446	0.253	0.802
	Tone×Object	-0.358	0.679	-0.528	0.600
(a) vs. (b)	Object	-0.085	0.400	-0.213	0.831
(c) vs. (d)	Object	0.293	0.414	0.707	0.480
3rd syllable	(Intercept)	-2.777	0.228	-12.201	
(2000–3000 ms)	Tone	0.878	0.322	2.724	<0.01***
	Object	0.736	0.522	1.411	0.170
	Tone×Object	0.941	0.793	1.186	0.242
(a) vs. (b)	Object	1.209	0.474	2.550	0.011 *
(c) vs. (d)	Object	0.261	0.672	0.388	0.701
4th syllable	(Intercept)	-1.615	0.203	-7.953	
(3000–4000 ms)	Tone	0.345	0.300	1.152	0.252
	Object	8.724	0.534	16.340	<0.01***
	Tone×Object	0.842	0.8322	1.011	0.317
(a) vs. (b)	Object	9.154	0.768	11.916	<0.01***
(c) vs. (d)	Object	8.300	0.596	13.937	<0.01***

. p=0.1, *p = 0.05, **p = 0.01, ***p = 0.001

5.7.3 Looks to [Target Single Object]

This section examines whether a surface T3 syllable has a predictive effect on a single structure interpretation similar to that of a T3S syllable on compound structure.

The analysis focuses on fixations on the [Target Single Object] in the four auditory stimuli conditions to clarify whether these fixations increase more rapidly when a surface T3 syllable is the input.

Figure 11 shows the proportion of fixations on the [Target Single Object] in the four auditory conditions. The upper panel shows the unambiguous T3 condition set whereas the bottom panel shows the ambiguous non-T3 condition set. The dashed blue lines represent the auditory single conditions while the solid red lines represent the auditory compound conditions.

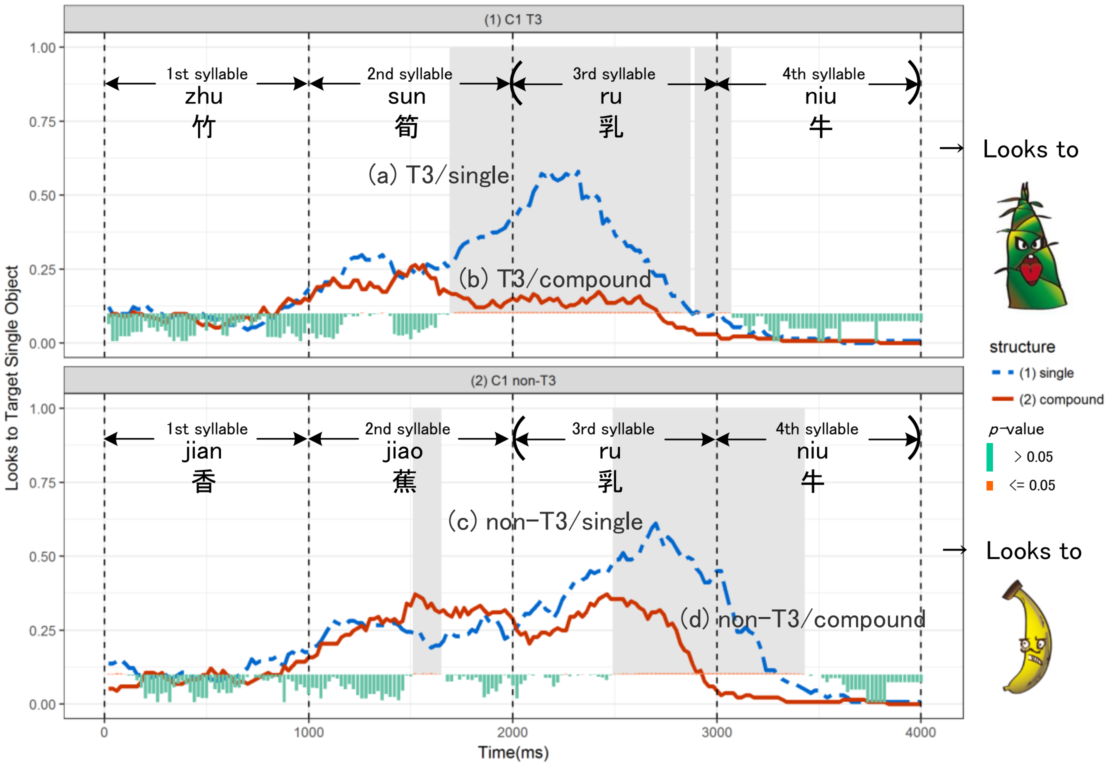


Figure 11. Proportion of fixations on [Target Single Object].

5.7.3.1 Permutation analysis

The permutation analysis showed that in the unambiguous T3 conditions, there were significantly fewer looks in the T3/compound condition (b) compared to the T3/single condition (a) in the 1700–2860 ms time window (cluster t-value = 289.011, cluster p-value < 0.001) and the 2900–3060 ms window (cluster t-value = 20.883, cluster p-value = 0.006). The 1700–2860 ms window started earlier than the onset of the head noun (1700 ms < 2000 ms). In the ambiguous non-T3 conditions, two significant clusters in the opposite direction appeared. The first is unexpected: there were fewer looks in the non-T3/single condition (c) compared to the non-T3/compound condition (d) at 1520–1640 ms (cluster *t*-value = 16.245, cluster *p*-value = 0.02). The second is consistent with the study's prediction: there were fewer looks in the non-T3/compound condition (d) compared to the non-T3/single condition (c) at 2500–3420 ms (cluster *t*-value = 219.293, cluster *p*-value < 0.001).

To examine whether a surface T3 syllable is more informative than a T3S syllable for identifying a target object, the permutation analysis was also used to confirm the time point of the divergence to the [Target Compound Object]. As **Figure 12** shows, in the unambiguous T3 conditions, there were significantly more looks to the [Target Compound Object] in the T3/compound condition (b) compared to the T3/single condition (a) in the 2060–4000 ms window (cluster *t*-value = 1009.457, cluster *p*-value < 0.001). In the unambiguous T3 conditions, the time point of divergence to the [Target Single Object] (1700 ms, **Figure 11**) was earlier than that to the [Target Compound Object] (2060 ms, **Figure 12**).

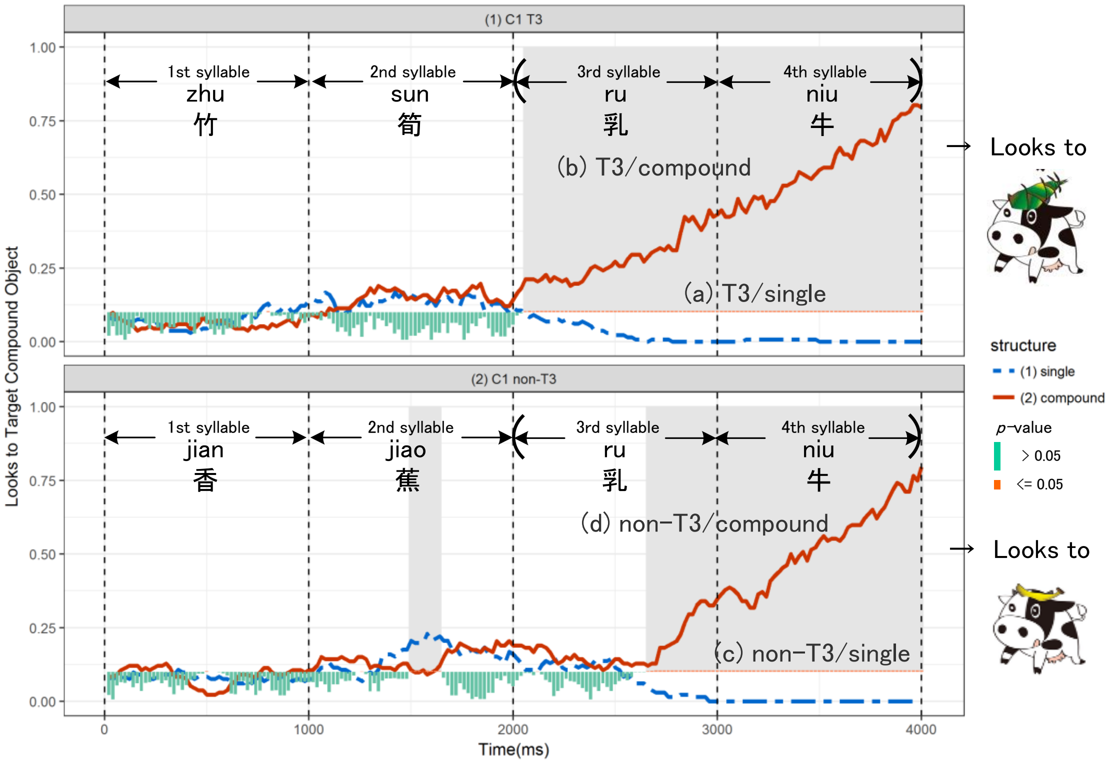


Figure 12. Proportion of fixations on [Target Compound Object].

5.7.3.2 LME analysis

The LME logit model results for every syllable interval (1000 ms) are reported in **Table 3**. The dependent variable was the empirical logarithm of looking against not-looking to the [Target Single Object]. Models included tone (T3 = 1; non-T3 = 0) and structure (single = 1; compound = 0) as predictor factors. The random slope structure included subject and item.

The fixations on the [Target Single Object] showed a pattern the reverse of that for the [Target Compound Object] and [Competitor Compound Object] combined (Section 5.7.1). At the 2nd syllable and the 3rd syllable, there were interaction effects between tone and structure, where the fixations on [Target Single Object] decreased most quickly in the T3/compound condition (b) whereas the fixations increased most quickly in the T3/single condition (a). While the non-T3/compound condition (d) also led to a lower amount of looks to the [Target Single Object] than in the non-T3/single condition (c), the degree of difference was smaller than that in the unambiguous T3 condition set.

Table 3

Summary of the Mixed Logit Model of Fixations on [Target Single Object]

		β	error	*t*	*p*
2nd syllable	(Intercept)	-1.912	0.240	-7.973	
(1000–2000 ms)	Tone	-0.354	0.353	-1.002	0.32
	Structure	0.196	0.298	0.657	0.51
	Tone×Structure	1.421	0.596	2.384	0.02*
(a) vs. (b)	Structure	0.895	0.412	2.171	0.03*
(c) vs. (d)	Structure	-0.512	0.431	-1.188	0.24
3rd syllable	(Intercept)	-0.671	0.245	-2.741	
(2000–3000 ms)	Tone	-0.158	0.351	-0.451	0.65
	Structure	5.860	0.297	19.710	<0.01***
	Tone×Structure	3.855	0.846	4.559	<0.01***
(a) vs. (b)	Structure	7.797	0.652	11.960	<0.01***
(c) vs. (d)	Structure	3.943	0.441	8.945	<0.01***
4th syllable	(Intercept)	-2.678	0.141	-19.001	
(3000–4000 ms)	Tone	-1.153	0.259	-4.452	<0.01***
	Structure	8.408	0.227	37.119	<0.01***
	Tone×Structure	-1.879	0.411	-4.575	<0.01***
(a) vs. (b)	Structure	7.470	0.180	41.500	<0.01***
(c) vs. (d)	Structure	9.345	0.387	24.137	<0.01***

. p=0.1, *$p = 0.05$, **$p = 0.01$, ***$p = 0.001$

5.8 Discussions

The current experiment led to some important findings about the function of T3S application in lexical processing.

5.8.1 Anticipatory effect of a T3S syllable on compound representation

The results presented in Section 5.7.1 suggest that listeners use the T3S input as a cue to anticipate a compound structure before encountering the head noun. That is, when listeners detect that the surface tonal representation (i.e., the T3S syllable) is different from the lexically determined tone (i.e., the underlying T3 syllable), they construct a compound representation (e.g., {bamboo-shoot +…}), which facilitates attention to any compound candidate that is consistent with the modifier input they have heard so far (e.g., {bamboo-shoot cow} or {bamboo-shoot raccoon}). Accordingly, listeners sent more looks to both the [Target Compound Object] and the [Competitor Compound Object] in the visual display when hearing the T3S syllable in the T3/compound condition.

The experiment further explored whether the predictive effect of a T3S syllable on compound structure relies on the tonal change in the modifier alone, and not on the segmental information of the head. As shown in **Figure 9**, the divergence in the unambiguous T3 condition set (1980 ms) occurs at a time point 20 ms earlier than the onset of the 3rd syllable (2000 ms). The 20 ms difference may not be robust evidence to support the anticipatory effect of a T3S syllable. However, previous research has found that there is a delay of approximately 150–200 ms for initiating eye saccade upon auditory input (Bahill & McDonald, 1983). If the initiation time for eye saccade is subtracted from the divergence time in the unambiguous T3 condition set, the actual processing to distinguish a compound structure from a single structure should start at around 1780 ms (i.e., 1980 ms minus 200 ms). That is, the predictive effect of the T3S syllable took place far earlier than the input of the head noun began. Therefore, the experimental results suggest that the anticipatory effect of a T3S syllable on the compound structure relies on the tonal change in the modifier alone; it does not wait for the head noun.

As discussed in Section 5.1, T3S application shows no strict one-to-one association between the tonal change and certain lexical structures. If the uncertainties of T3S application cancel out the predictive effect of T3S, listeners should not be able to distinguish an upcoming compound structure from an upcoming single structure when they encounter the T3S syllable input. However, this experiment showed a faster increase in looks to compound objects in the T3S condition, which started earlier than the head noun input. This finding suggests that the predictive effect of a T3S syllable is robust regardless of the uncertainties of T3S application. The strong preference for pronouncing a T3S syllable in a compound noun may play an important role in a T3S syllable's effect on predictions of upcoming lexical structure.

In contrast to the unambiguous T3 conditions, the two non-T3 conditions show no obvious tonal difference in their first noun, which should lead to temporary ambiguity between the single and compound structure. The relatively small difference that was observed at the 3rd syllable between the two non-T3 conditions indicates that listeners cannot distinguish a non-T3/compound condition from a non-T3/single condition until they have detected the segmental information of the head noun.

5.8.2 No effect of a T3S syllable on computing the meaning of the compound

The prediction that a T3S syllable would have an accelerating effect, as the CAR does, on the listeners' computation of the meaning of a compound based on the previous anticipation of the compound representation was not borne out. If the prediction were true, there should have been a rapid increase in looks to the "correct answer" [Target Compound Object] along with a rapid decrease of looks to the "wrong answer" [Competitor Compound Object] when participants were hearing the T3/compound condition.

As described in Section 5.7.2, the divergence was earlier in the unambiguous T3 condition set than in the ambiguous non-T3 condition set. However, the difference was only 100 ms, and there was no interaction effect, but only a main effect of object, in the divergence window (both in the 2700–4000 ms interval and the 4th syllable window). This result indicates that the tone type of the modifier may not affect the identification of the compound target. In other words, the experimental results do not provide evidence that the computation of meaning in the T3S compound condition is faster than in the non-T3 compound condition. This tendency is not consistent with the findings on the effects of the Japanese CAR reported by Hirose and Mazuka (2015). One possible reason is that when the CAR applies, the following constituent must be the head of a compound, while when T3S applies, various structures can follow. Therefore, a Mandarin listener may need more time to identify the head noun (i.e., to eliminate other possibilities) than did the

Japanese listeners in Hirose and Mazuka’s (2015, 2017) studies.

It is unclear why some unexpected effects in the opposite direction were observed in the 420–600 ms and 2360–2480 ms windows in the non-T3 conditions. As discussed in Section 1.2, the duration of the vowel affects syllabification; some acoustic properties such as vowel duration in the non-T3 conditions may affect the prediction of lexical structure. However, because the effects had relatively small *t*-values and only occurred in the non-T3 conditions at a time point earlier than the divergence onset in the unambiguous T3 conditions, they should be independent from a T3S effect. These effects are therefore left out of further discussions.

5.8.3 Anticipatory effect of a surface T3 syllable on a single word representation

The results presented in Section 5.7.3 suggest that, in addition to the anticipatory effect of a T3S syllable on compound structure, the information of a surface T3 syllable also helps listeners to anticipate a single structure interpretation. The interactions found at the 2nd and the 3rd syllables show that when listeners heard a surface T3 syllable in the T3/single condition, they rapidly anticipated a single noun head. Furthermore, the divergence in the T3 conditions to the [Target Single Object] (1700 ms, **Figure 11**) occurs at an earlier time point than the divergence to the [Target Compound Object] + [Competitor Compound Object] (1980 ms, **Figure 9**) as well as that to the [Target Compound Object] (2060 ms, **Figure 12**). These results suggest that the anticipatory effect of a surface T3 syllable on a single word reading starts even earlier than the anticipatory effect of a T3S syllable on a compound word reading.

The anticipatory effect of the surface T3 syllable is robust; however, it is still unknown why an unmarked T3 syllable has an earlier effect than a marked T3S syllable. One possible reason may be due to the experimental context. Each visual display included one single object but two compound objects corresponding to the auditory stimulus. For this reason, a surface T3 syllable in a single condition may have been more informative than a T3S syllable in a compound condition.

In contrast to the unambiguous T3 conditions, in the ambiguous non-T3 conditions the differentiation started after the onset of the head noun. This is because there was no tonal difference between the two non-T3 conditions at the first noun. The listeners had to wait until they encountered the head and then disambiguate the non-T3/compound condition from the non-T3/single condition.

An unpredicted effect in the opposite direction was observed between the two non-T3 conditions in the 1520–1640 ms window. The cause is still unclear, but it may be related to some acoustic properties of non-T3 syllables, as mentioned, such as vowel duration (see Section 5.8.2). However, this effect in the non-T3 conditions had a comparatively small *t*-value and occurred earlier than the divergence onset in the unambiguous T3 conditions, suggesting that the effect is isolated from the T3S effect; it is a topic for further examination.

5.9 Conclusions

Experiment 1 found, first, that Mandarin listeners are capable of using the preference for a T3S syllable in compound nouns as a prosodic cue to anticipate lexical structure in online processing. In the restricted experimental environment that only included the options of single nouns or compound nouns, listeners used the help of the T3S information to anticipate a compound structure and eliminate a single structure as a possibility before they encountered the segmental information of the head noun. The results also suggest that the robustly predictive effect of T3S is not affected by the uncertainties of T3S application. However, this experiment did not provide further evidence that input of a T3S syllable facilitates the computation of compound nouns.

Chapter 6

Experiment 2: The anticipatory effect of T3S on tone type

As discussed in Chapter 1, the necessary conditions of a T3S syllable are (i) that another syllable follow it and (ii) that the following syllable is of the underlying T3 type. These factors can be considered at two different levels: the former is related to structural representation while the latter is related to the tonal information stored in a lexical entry. Experiment 1 (Chapter 5) demonstrated that Mandarin listeners are able to use the presence of a T3S syllable as a cue to anticipate compound structure, which is related to the first necessary condition; that is, that there be a following syllable. In the current chapter, Experiment 2 explores the second necessary condition of a T3S syllable; specifically, the experiment tests whether the T3S syllable is used to predict the tone type of the following syllable.

6.1 Introduction

The phonological form of current input helps listeners to estimate relative phonological likelihoods for upcoming words. In English, for example, indefinite articles obey a phonological regularity that *an* is followed by a noun that begins with a vowel sound (e.g., *an* + *airplane*) while *a* is followed by a noun beginning with a consonant (e.g., *a* + *kite*). Using an event-related brain potential experiment (ERP), DeLong, Urbach, and Kutas (2005) found that after listeners heard a sentence such as *The day was breezy so the boy went outside to fly…*, a following indefinite article *an* evoked a increased N400 effect compared to a following *a*, and a vowel-initial noun such as *airplane* following the indefinite article again led to a increased N400 effect compared to the consonant-initial noun *kite*. These results suggest that listeners can pre-activate semantic features of categories depending on the preceding context (e.g., the high possibility of a boy on a breezy day flying a kite), and also anticipate various syntactic aspects of to-be-presented material (e.g., by associating a consonant-initial noun such as *kite* with the indefinite article *a*).

The tone types of Mandarin Chinese words are lexically determined (e.g., *guan3* 'plumber' [管], *guan4* 'jug' [罐]). Given the case that these words follow an underlying T3 word (e.g., *shui3* 'water' [水]), the preceding underlying T3 word may change its tone type while the following word always keeps its original tone type (e.g., *shui3* + *guan3* → *<u>shui2</u>* *guan3* 'water plumber' [水管], *shui3* + *guan4* → *shui3 guan4* 'water jug' [水罐]). If Mandarin listeners are able to use this association in online processing, and further activate a lexical item with the predicted tone type, it is possible that a T3S syllable as input helps listeners anticipate a T3 syllable to follow. On the other hand, input of a surface T3 syllable should facilitate a following word of a non-T3 type. Taking an extreme case for an example, when listeners must choose the following word from two candidates, *guan3* 'plumber' and *guan4* 'jug' after the input of *shui* 'water', they may be expected to rapidly choose *guan3* when *shui* 'water' is pronounced as a T3S syllable of *shui2*, or to rapidly choose *guan4* when *shui* 'water' is pronounced with its underlying tone type, as *shui3*.

There are only a very few cases in Mandarin Chinese where a T3 word can be followed by either of two homophones with different tones (e.g., *shui3* + *guan3* vs. *shui3* + *guan4*). However, if the anticipatory effect of a T3S syllable on tone type is robust during lexical processing, a group of words sharing the same tonal feature (e.g., the same onset of T3 or non-T3) should be activated as upcoming candidates on the basis of the tone type of the input so far. That is, given an environment in which exist multiple candidates of T3 or non-T3 words, an input of the T3S syllable *shui2* 'water' should active a group of T3 words (e.g., *jiao3* 'dumpling' [餃], *cao3* 'grass' [草]) while an input of the surface T3 syllable *shui3* 'water' should active a group of non-T3 words (e.g., *bei1* 'cup' [杯], *ku4* 'dam' [庫]). This chapter reports on a visual world paradigm experiment using multiple candidates with T3 or non-T3 head nouns in the same visual display to clarify whether tonal change in the modifier affects lexical selection in online processing.

6.2 Materials

The materials of Experiment 2's visual world paradigm included auditory stimuli and the corresponding visual displays. The lexical pattern used in the auditory stimuli and visual displays were all noun + noun compounds. The first noun represents a plant as the modifier of the compound (e.g., *zhu2 sun3* 'bamboo-shoot', *xiang1 jiao1* 'banana'), and the second noun represents an animal as the head of the compound (e.g., *ru3 niu2* 'cow', *hu2 die2* 'butterfly').

6.2.1 Auditory Stimuli

The auditory stimuli consisted of noun + noun compounds in which each noun was of two syllables. The stimuli manipulated the underlying tone type, as T3 or non-T3, of (i) the modifier noun and (ii) the head noun. For example, in the first case, the modifier *zhu2* <u>*sun3*</u> 'bamboo-shoot' is a "T3 modifier" because of the tone of its second syllable *sun3*, while the modifier *xiang1 jiao1* 'banana' is a "non-T3 modifier" due to the tone of its second syllable *jiao1*. For the heads, *ru3* in *ru3 niu2* 'cow' is an example of a "T3 head", while *hu2* in *hu2 die2* 'butterfly' is a "non-T3 head." These manipulations generated the experiment's four auditory conditions: (a) T3S + T3, (b) T3 + non-T3, (c) non-T3 + T3, and (d) non-T3 + non-T3. The conditions are exemplified in (20). Only in the T3S + T3 condition (a) is the T3S rule applied. Its second syllable was pronounced as surface T2 in the auditory stimuli (e.g., *zhu2 sun3→2 ru3 niu2*). In contrast, the second syllable in the T3 + non-T3 condition (b) remains T3 because it is followed by a non-T3 head (e.g., *zhu2 sun3 hu2 die2*). The non-T3 + T3 condition ([c], e.g., *xiang1 jiao1 ru3 niu2*) and the non-T3 + T3 condition ([d], e.g., *xiang1 jiao1 hu2 die2*) do not carry any information related to T3S application at the second syllable position as their modifiers are non-T3 types. They were used as baseline conditions for comparison to the T3S + T3 condition (a) and the T3 + non-T3 condition (b), respectively.

(20)

a. T3S + T3
[*zhu2* 竹 *sun3* 筍] [*ru3* 乳 *niu2* 牛]→ 2-2-3-2 'bamboo-shoot cow'
b. T3 + non-T3
[*zhu2* 竹 *sun3* 筍] [*hu2* 蝴 *die2* 蝶] 'bamboo-shoot butterfly'
c. non-T3 + T3
[*xiang1* 香 *jiao1* 蕉] [*ru3* 乳 *niu2* 牛] 'banana cow'
d. non-T3 + non-T3
[*xiang1* 香 *jiao1* 蕉] [*hu2* 蝴 *die2* 蝶] 'banana butterfly'

The auditory stimuli were recorded using the same procedure as in Experiment 1 (the auditory stimuli used in all experiments in this dissertation were recorded at the same time). The auditory stimuli for the T3S + T3 (a) and non-T3 + T3 (c) conditions for the current experiment were recycled from Experiment 1. All the stimuli were pronounced at a speed of approximately one syllable/1000 ms with natural pauses (based on a stopwatch).

6.2.2 Visual stimuli

Each display contained four objects with compound noun names. **Figure 13** shows the example of the visual display used for the auditory stimulus *zhu2 sun2 ru3 niu2* 'bamboo-shoot cow' (竹筍乳牛) of the T3S + T3 condition (a). All four objects are depicted so that they share the same modifier {bamboo-shoot…}.[19] Two of these objects' names have T3 heads while the other two have non-T3 heads. To give a concrete example, as illustrated in **Figure 13**, the right-bottom object {bamboo-shoot cow} (*zhu2 sun3+ ru3 niu2*) is the [Target Object], which was used as the final correct answer to the auditory stimulus. The left-bottom object {bamboo-shoot raccoon} (*zhu2 sun3 + wan3**[20] *xiong2* [竹筍浣熊]) is the [Competitor Object]. The head noun of the [Competitor Object] had the same tone type but a different referent than the [Target Object] (e.g., the head of the [Competitor Object] is {raccoon} while that of the [Target Object] is {cow}; however, both {raccoon} (*wan3 xiong2*) and {cow} (*ru3 niu2*) are of the T3 type). The upper two objects {bamboo-shoot butterfly} (*zhu2 sun3 + hu2 die2* [竹筍蝴蝶]) and {bamboo-shoot penguin} (*zhu2 sun3 + qi4* e2* [竹筍企鵝]) are [Distractor] objects, which have heads of different tone types from the [Target Object] (e.g., *hu2 die2* and *qi4 e2* are of the non-T3 type).

Figure 13. An example of a visual display used in Experiment 2.

The experimental design allows the examination of whether listeners can anticipatorily limit the range of candidate objects at an early stage if any tonal change such as a T3S syllable is provided in the modifier, or if they must rely on the segmental information of the head noun to identify the referent. Taking **Figure 13** as an example, if listeners use the tonal change in a T3S modifier as a cue to predict an upcoming T3 head, they should active the two objects with a T3 head: the [Target Object] (*zhu2 sun3 + ru3 niu2*) and the [Competitor Object] (*zhu2 sun3 + wan3 xiong2*), while excluding the two

---

[19] The plant or animal names in curly brackets represent the visual objects. The tonal description after the curly brackets represents the underlying tone but not the actual pronunciation.

[20] A syllable marked by an asterisk (*) has different pronunciations in Taiwan Mandarin and Beijing Mandarin. Refer to **Appendix C** for all such items used in Experiment 2.

[Distractor] objects with non-T3 heads (*zhu2 sun3 + hu2 die2* and *zhu2 sun3 + qi4 e2*). In contrast, if listeners rely on the segmental information of the head noun to identify the correct visual object, without taking the tone type of these heads into consideration, the [Target Object] should be the only target the listeners search for, and hence there should be no difference between the number of participants' eye fixations on the [Competitor Object] and the number of fixations on the two [Distractor] objects. The number of fixations on the [Competitor Object] was employed as an important indicator for examining whether the tonal information has predictive effects.

The visual displays used for the T3S + T3 condition (a), such as **Figure 13**, were also used for the T3 + non-T3 condition (b). For example, in the case of the auditory stimulus *zhu2 sun3 hu2 die2* 'bamboo-shoot butterfly', the upper-left object {bamboo-shoot butterfly} (*zhu2 sun3 + hu2 die2*) was used as the [Target Object]. The upper-right object {bamboo-shoot penguin} (*zhu2 sun3 + qi4 e2*) was the [Competitor Object]. The two bottom objects, {bamboo-shoot cow} (*zhu2 sun*3 + *ru3 niu2*) and {bamboo-shot raccoon} (*zhu2 sun3 + wan3 xiong2*), were [Distractor] objects. Again, if surface T3 in a T3 modifier is also used to predict a non-T3 head, both the [Target Object] and the [Competitor Object] should be activated at an early stage. If not, listeners should only search for the [Target Object] but not differentiate the [Competitor Object] from [Distractor] objects.

The visual displays used for the non-T3 conditions, non-T3 + T3 (c) and non-T3 + non-T3 (d), were constructed independently from those for the T3 conditions. Taking the auditory stimuli in (20c) and (20d) as an example, the corresponding visual display was made up of four objects modified by bananas. Two of them had T3 heads ('banana cow', *xiang1 jiao1 + ru3 niu2* [香蕉乳牛] and 'banana raccoon', *xiang1 jiao1 + wan3 xiong2* [香蕉浣熊]). The other two objects had non-T3 heads ('banana butterfly', *xiang1 jiao1 + hu2 die2* [香蕉蝴蝶] and 'banana penguin', *xiang1 jiao1 + qi4 e2* [香蕉企鵝]). It is reasonable to presume that listeners would not differentiate a [Competitor Object] from a [Distractor] in both non-T3 conditions, such as (20c) and (20d), at an early stage, because there is no restriction from a non-T3 modifier on the tone type of its following head. Therefore, the non-T3 conditions provided baseline observations for comparison to the T3 conditions.

6.2.3 Filler items

There were two patterns of visual display used for the auditory filler items. One pattern contains all T3 + T3 compounds (e.g., *zhu2 sun3 + ru3 niu2* 'bamboo-shoot cow') for the four visual objects in a display, while the other pattern contains all non-T3 + non-T3 compounds (e.g., *xiang1 jiao1* + *wu1 guei1* ‘banana turtle’). Hence, a listener could not obtain any information related to the distinction between T3 and non-T3 heads from the visual context.

The experiment employed 24 experimental items and 24 filler items. The experimental items were divided into four lists using a Latin-square design, and the filler items were used in common in each list. Each list for one participant included 48 trials.

6.3 Participants

The same participants of Experiment 1 took part in Experiment 2.

6.4 Procedure

The procedure was the same as that of Experiment 1. Eye movements were recorded from the onset of auditory stimuli until the participant clicked the mouse (**Figure 14**).

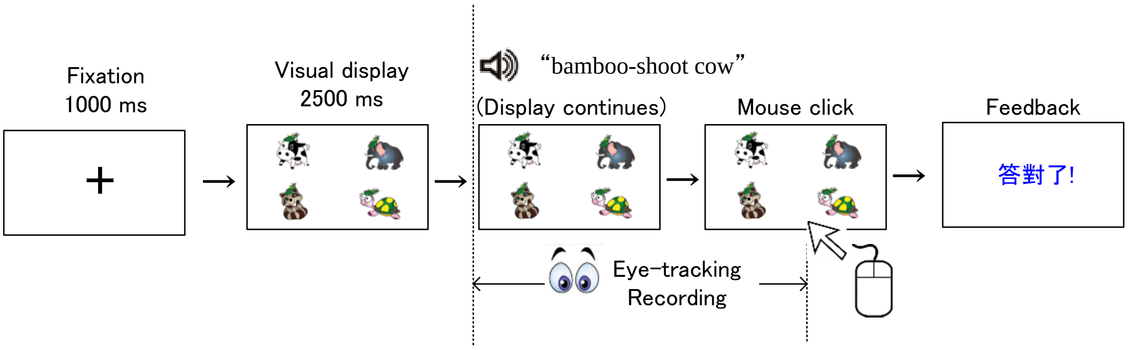


Figure 14. The procedure used in Experiment 2.

6.5 Data analysis

The analysis methods were the same as those employed in Experiment 1. Permutation cluster analysis was used to observe the explicit time interval of the differences in fixation proportions between the T3 head conditions and the baseline non-T3 head conditions; that is, comparisons were made between the T3S + T3 condition (a) and the non-T3 + T3 condition (c), and between the T3 + non-T3 condition (b) and the non-T3 + non-T3 condition (d). The dependent variable in the permutation analysis was the proportion of fixations by subject in every 20 ms time window. The permutation distribution of *t*-values was obtained by resampling 1,000 times the labels of each paired condition on the original data points within the time window clusters that showed a significant difference.

In addition, the study used LME analysis to evaluate interactive effects in each syllabic interval (every 1000 ms). The dependent variable for the LME analysis was the log odds (empirical logit) of the total fixation time of looking against not-looking to AOI(s) (i.e., to certain objects or combinations of two objects). The LME models included modifier (T3 = 1, non-T3 = 0) and head (T3 = 1, non-T3 = 0) as predictor factors. The variable of each predictor factor was scaled and centered. The random slope structure included both subject and item.

6.6 Predictions

There is no question that the listener will eventually look at the [Target Object] in all conditions. However, if the presence of a T3S syllable has predictive effect on the tone type of the head, more fixations on both the [Target Object] and the [Competitor Object] of a T3 head should occur earlier in the T3S + T3 condition (a) compared to the non-T3 + T3 condition (c). If the presence of a surface T3 syllable has the same predictive effect as well, more fixations on the [Target Object] and [Competitor Object] of a non-T3 head should occur earlier in the T3 + non-T3 condition (b) compared to the non-T3 + non-T3 condition (d). Furthermore, if the predictive effect of a T3S or surface T3 syllable relies only on the tonal feature of the modifier, the temporary divergence between T3 conditions and non-T3 conditions should start before the listener encounters the head noun.

More concretely, if the study's assumption that T3S and underlying T3 syllables have a predictive effect, the following situations should be observed in the experiment. In the first case, only the tonal change in a T3S syllable gives listeners a processing cue because the T3S rule is allowed not to apply in a T3S environment and therefore a surface T3 syllable is ambiguous regarding whether it will be followed by a T3 word or a non-T3 word. In this situation, that is, if a T3S syllable is more informative than a surface T3 syllable, the experimental results should show an interaction effect in the 2nd syllable where the difference between the T3S + T3 condition (a) and the non-T3 + T3 condition (c) is larger than that between the T3 + non-T3 condition (b) and the non-T3 + non-T3 condition (d), and the onset of the divergence in the T3 head conditions will start earlier than that in the non-T3 conditions.

The second situation is that a T3S syllable and a surface T3 syllable carry the same amount of information. That is, listeners will use two associations, one between the T3S syllable and the following T3 word and one between a surface T3 syllable and a following non-T3 word. In this case, the experimental results will show a main effect of the tone type of the modifier from the 2nd syllable, which should be larger in the T3S + T3 condition (a) than in the non-T3 + T3 condition (c), and at the same time should also be larger in the T3 + non-T3 condition (b) than in the non-T3 + non-T3 condition (d). The divergence onset in these two condition sets should be the same.

6.7 Results

This section reports two sets of analyses, first, of the fixations on the [Target Object] and the [Competitor Object] combined to examine the anticipatory effect on the tone type of the upcoming head and observe when the effects appear. The second analysis is of fixations on the [Target Object] to confirm whether the previous anticipation of tone type facilitates the final identification of the correct target.

6.7.1 Looks to [Target Object] + [Competitor Object]

Figure 15 shows the proportion of fixations on the [Target Object] and [Competitor Object] combined in the four auditory conditions. The four auditory conditions are split into two panels by the tone type of the head noun: the upper panel presents both T3 head conditions, T3S + T3 (a) versus non-T3 + T3 (c), while the bottom panel presents the non-T3 head conditions, T3 + non-T3 (b) versus non-T3 + non-T3 (d). The colors of the lines represent the tone type of the modifier: the solid red lines are for the T3 modifier while the dashed blue lines are for the non-T3 modifier. The gray shading shows the time spans where a significant difference was identified by the permutation analysis.

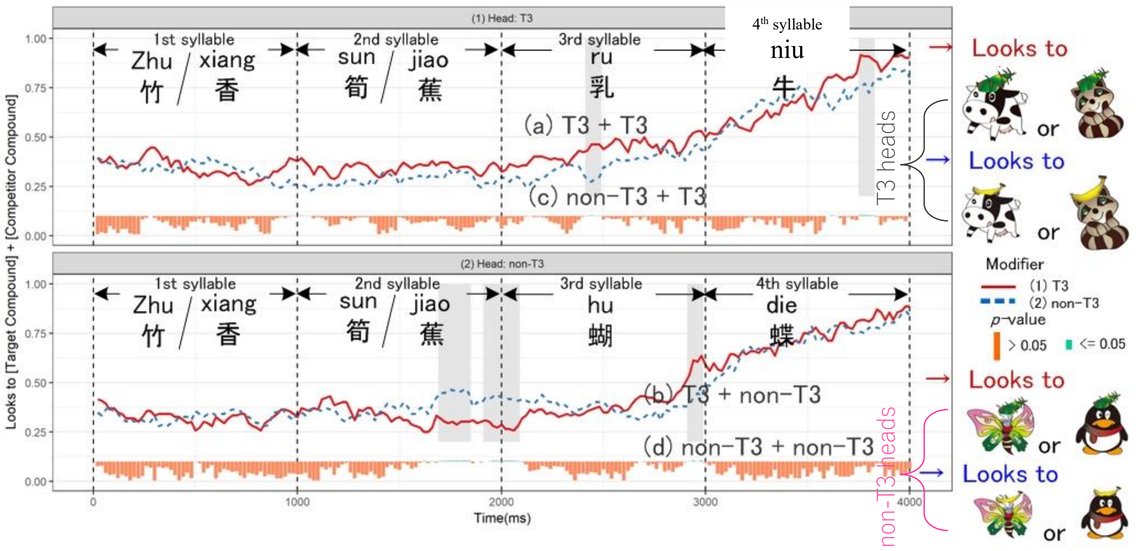


Figure 15. Proportion of fixations on [Target Object] + [Competitor Object].

6.7.1.1 Permutation analysis

The permutation analysis showed that, in the T3 head conditions, the number of fixations on the [Target Object] and [Competitor Object] combined was higher in the T3S + T3 condition (a) compared to the non-T3 + T3 condition (c) during the 2420–2480 ms window (cluster *t*-value = 8.908, cluster *p*-value < 0.001) and the 3760–3820 ms window (cluster *t*-value = 12.646, cluster *p*-value < 0.001). In the non-T3 head conditions, significant differences were found in three intervals. The first two intervals were unexpected. There were more looks to the baseline non-T3 + non-T3 condition (d) than to the T3 + non-T3 condition (b) in the 1700–1840 ms window (cluster *t*-value = 23.893, cluster *p*-value < 0.001) and the 1920–2080 ms window (cluster *t*-value = 24.473, cluster *p*-value < 0.001). The third interval was in the expected direction, with higher numbers of fixations in the T3 + non-T3 condition (b) than in the baseline non-T3 + non-T3 condition (d) in the 2920–2980 ms window (cluster *t*-value = 10.245, cluster *p*-value = 0.001).

6.7.1.2 LME analysis

The results of the LME modeling are reported in **Table 4**. In the 2nd syllable, the model showed an interaction effect between modifier and head. The direction of the effect was inconsistent with what was expected. It reflected an advantage of the baseline non-T3 + non-T3 condition (d) over the T3 + T3 condition (b) while there was no difference between the T3S + T3 condition (a) and the non-T3 + T3 condition (c). In the 3rd syllable, there was a marginal main effect of modifier, consistent with the study's assumption. It was due to the marginally higher number of fixations in the T3S + T3 condition (a) compared to the baseline non-T3 + T3 condition (c). The 4th syllable also shows a main effect of modifier but it was due to higher numbers of fixations in the T3 + non-T3 condition (b) than in the baseline non-T3 + non-T3 condition (d).

Table 4

Summary of the Mixed Logit Model of Fixations on [Target Compound] + [Target Single] Objects

		β	*SE*	*t*	*p*
2nd syllable (1000–2000 ms)	(Intercept)	-0.156	0.226	-0.690	0.494
	Modifier	0.178	0.451	0.393	0.696
	Head	-0.606	0.422	-1.436	0.152
	Modifier×Head	1.691	0.844	2.003	0.046 *
(a) vs. (c)	Modifier	1.001	0.663	1.511	0.138
(b) vs. (d)	Modifier	-0.675	0.657	-1.029	0.309
(a) vs. (b)	Head	0.242	0.599	0.404	0.687
(c) vs. (d)	Head	-1.439	0.590	-2.440	0.016 *
3rd syllable (2000–3000 ms)	(Intercept)	0.449	0.235	1.913	0.070 .
	Modifier	0.601	0.336	1.792	0.074 .
	Head	0.098	0.336	0.291	0.771
	Modifier×Head	0.507	0.671	0.756	0.450
(a) vs. (c)	Modifier	0.853	0.481	1.775	0.077 .
(b) vs. (d)	Modifier	0.350	0.512	0.683	0.498
4th syllable (3000–4000 ms)	(Intercept)	5.130	0.265	19.391	<0.001***
	Modifier	0.682	0.331	2.059	0.045 *
	Head	-0.411	0.261	-1.577	0.116
	Modifier×Head	-0.385	0.521	-0.739	0.460
(a) vs. (c)	Modifier	0.488	0.439	1.111	0.272
(b) vs. (d)	Modifier	0.872	0.483	1.808	0.078 .

. p=0.1, *p = 0.05, **p = 0.01, ***p = 0.001

6.7.2 Looks to [Target Object]

Figure 16 shows the proportion of fixations on the [Target Object] in the four auditory conditions. The layout is the same as in **Figure 15**.

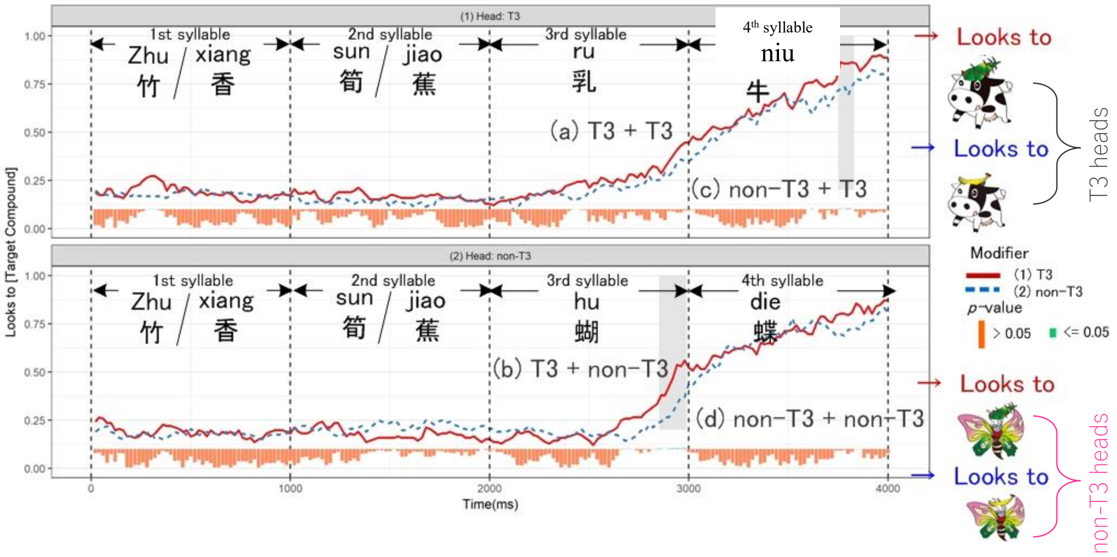


Figure 16. Proportion of fixations on [Target Compound] objects.

6.7.2.1 Permutation analysis

Areas of significant difference in the T3 head conditions were observed in the 3760–3820 ms interval (cluster *t*-value = 10, cluster *p*-value < 0.001) where the T3S + T3 condition (a) showed a higher number of fixations than did the baseline non-T3 + T3 condition (c). On the other hand, a significant difference was also found in the non-T3 head conditions in the 2860–2980 ms window (cluster *t*-value = 19.317, cluster *p*-value < 0.001), where the T3 + non-T3 condition (b) had a higher number of fixations than the baseline non-T3 + non-T3 condition (d).

6.7.2.2 LME analysis

The results of the LME models for each syllable are summarized in **Table 5**. No effect was found in the 2nd syllable. In the 3rd syllable, a main effect of modifier is seen, which was due to the marginally higher number of fixations in the T3S + T3 condition (a) compared to the baseline T3 + non-T3 condition (c) as well as to the marginally higher number of fixations in the T3 + non-T3 condition (b) compared to the baseline non-T3 + non-T3 condition (d). In the 4th syllable, there was a main effect of modifier and a marginal main effect of head, which was due to a significantly higher number of fixations in the T3S + T3 condition (a) than in the non-T3 + T3 condition (c), while there was no difference between the T3 + non-T3 condition (b) and the non-T3 + non-T3 condition (d).

Table 5

Summary of the Mixed Logit Model of Fixations on [Target Compound] Objects

		β	error	*t*	*p*
2nd syllable (1000–2000 ms)	(Intercept)	-3.150	0.258	-12.192	<0.001***
	Modifier	-0.159	0.517	-0.308	0.759
	Head	-0.479	0.370	-1.295	0.196
	Modifier×Head	0.873	0.739	1.181	0.238
(a) vs. (c)	Modifier	0.265	0.6244	0.425	0.673
(b) vs. (d)	Modifier	-0.608	0.6718	-0.905	0.370
3rd syllable (2000–3000 ms)	(Intercept)	-1.959	0.190	-10.287	<0.001***
	Modifier	0.847	0.349	2.422	0.020 *
	Head	-0.012	0.336	-0.035	0.971
	Modifier×Head	-0.087	0.671	-0.130	0.897
(a) vs. (c)	Modifier	0.802	0.475	1.689	0.098 .
(b) vs. (d)	Modifier	0.887	0.496	1.789	0.080 .
4th syllable (3000–4000 ms)	(Intercept)	4.269	0.331	12.895	<0.001***
	Modifier	1.066	0.360	2.958	0.005 **
	Head	-0.560	0.293	-1.912	0.056 .
	Modifier×Head	0.419	0.586	0.715	0.475
(a) vs. (c)	Modifier	1.280	0.493	2.594	0.013 *
(b) vs. (d)	Modifier	0.865	0.524	1.651	0.106

. p=0.1, *p = 0.05, **p = 0.01, ***p = 0.001

6.8 Discussions

6.8.1 No robust effect on head tone type prediction

As discussed in Section 6.6, Experiment 2 assumed two possible situations in which a T3S facilitates the listener's anticipation of the tone type of the head. The first situation is that if a T3S syllable is more informative than a surface T3 syllable, we should observe an interaction effect from the 2^{nd} syllable. The other situation is that if a T3S syllable and a surface T3 syllable carry the same amount of information, we should observe a main effect of modifier from the 2nd syllable. However, the experiment's results do not provide evidence to support either assumption.

First, it is unclear whether the effect observed in the 1700–1840 ms and 1920–2080 ms intervals indicated that a non-T3 modifier was readily associated with a non-T3 head. However, the more expected effect of the T3S syllable was observed from 2420 ms. Supposing that a T3S syllable is more informative for the choice of the following head, it is conceivable that the effects observed in the 1700–1840 ms window and the 1920–2080 ms window in the non-T3 head conditions could be unrelated to the effect of the T3S syllable, but derive instead from some acoustic feature of the non-T3 syllable.

In the 3rd and the 4th syllable, the findings did show a main effect (and marginal main effect) of modifier in the T3S and the surface T3 conditions. The direction of the effect is consistent with what we would expect if a T3S syllable and a surface T3 syllable are equally informative. However, the tendency shown in the post hoc analysis was not robust, and the transient effects identified by the permutation analysis were very short (the three significant intervals were all 60 ms).

6.8.2 No robust effect on target compound object identification

It is reasonable to assume that if the tonal feature of the modifier helps listeners to anticipate the tone type of the head, the anticipation should further facilitate the computation of the target compound. For that, the fixations on the [Target Object] (**Figure 16**) should exhibit a similar tendency as the fixations on the [Target Object] and [Competitor Object] combined (**Figure 15**). That is, a temporary advantage would be seen in the T3 modifier conditions compared to the baseline non-T3 modifier conditions. Furthermore, because word identification involves the computation of a head noun, the divergence onsets in fixations on the [Target Object] (**Figure 16**) should start later than the onset of the head noun.

However, the analysis found no main effect of modifier in either the 3rd or 4th syllable; the post hoc analysis showed no robust difference, and the effect intervals identified by the permutation analysis were very short (60 ms in the T3 head condition set, 120 ms in the non-T3 head condition set). For the same reasons mentioned in Section 6.8.2, these results may be insufficient to support the facilitative effect of T3S or T3 syllables in the identification of a target compound object.

6.8.3 No evidence for a predictive effect on phonological form

The experiment did not provide robust evidence that Mandarin listeners use T3S as a prosodic cue to predict the tone type of the following head.

There are two possible explanations for the lack of evidence. The first is that people do not use T3S as a cue to predict the tone type of upcoming material. Therefore, even though there are some lexical item candidates whose tone types match the condition that follows a T3S or surface T3 syllable, an input of a T3S or surface T3 syllable does not activate these lexical item candidates.

Another possibility is that people in fact are able to predict phonological form, but that the retrieval of the lexical entries for visual objects' names takes too much time. Hence, the predictive effect on phonological form would occur after, or be cancelled by, the identification of the head noun. As reviewed in Section 2.4.2, Ito, Pickering, and Corley (2018) tested both English native speakers (L1) and Japanese English learners (L2) to investigate predictive effects on phonological form. They found a predictive effect in L2 as well as in L1; however, the effect in L2 occurs 600 ms later than the target word onset. They considered that the time lag may have been due to the time the L2 speakers needed to retrieve names when seeing pictures. If this was the case in the current study's Experiment 2, it might have been a burden for listeners to retrieve all the names of the four objects in the visual displays, and then to extract the phonological information from the lexical entries. The situation is different from that of Experiment 1, where the listeners could visually distinguish a complicated object, which corresponded to a compound structure, from a simpler object, which corresponded to a single structure. If the lack of effect in Experiment 2 is caused by the complexity of lexical retrieval, the predictive effect may yet be confirmed if the burden of the task can be reduced. For instance, the four visual objects could be substituted by orthographic representations (e.g., Mandarin characters) or the number of visual objects in a display could be reduced from four to two to shorten the retrieval time needed.

6.9 Conclusion

The main purpose of Experiment 2 was to examine whether T3S information helps listeners predict the tone type of following syllables. Despite showing some transient effects, the experiment's results were insufficient to support a predictive effect of T3S syllables on the tone type of following components. Taking the results of Experiment 1 and Experiment 2 together, they imply that Mandarin listeners do not use all sources of prosodic information related to T3S to predict the features of upcoming components. They anticipate structural representation but do not predict tone type (or associated meanings) in online lexical processing.

Chapter 7

Conclusion

7.1 The critical questions

This dissertation has investigated the predictive effect of Mandarin tone 3 sandhi in lexical processing. It aimed to answer two critical research questions, repeated here as (21), which are based on the two necessary conditions of the T3S rule (T3→T2/ _T3): first, that the T3S syllable must be followed by another syllable, and second, that the following syllable must be of the T3 type.

(21)
Research Question 1: Does T3S help listeners to predict lexical structure?
Research Question 2: Does T3S help listeners to predict phonological form?

The study employed two visual world paradigm experiments to address these questions. Experiment 1 (Chapter 5) found that input of a T3S syllable (e.g., *zhu2 sun3→2…* 'bamboo-shoot') helped listeners predict a compound structure (e.g., [[*zhu2 sun3*]$_{N}$ […]$_{N}$]$_{NP}$, such as *zhu2 sun3* 'bamboo-shoot'+ *ru3 niu2* 'cow'), while input of a surface T3 syllable (e.g., *zhu2 sun3* 'bamboo-shoot') helped listeners predict a single structure (e.g., [*zhu2 sun3*]$_{NP}$). These effects occurred before the listeners encountered the head noun of the compound (i.e., before the ambiguity between the single and the compound structure was resolved by further input). The results of Experiment 1 provided an answer to the first critical question: Mandarin listeners are capable of using T3S as a prosodic cue to anticipate lexical structure, and the anticipation relies solely on the tonal information in the T3S syllable itself (i.e., it does not rely on the segmental information of the head noun).

On the other hand, Experiment 2 (Chapter 6) found no predictive effect of a T3S syllable on upcoming tone type. This experiment considered the possibility that if T3S does help listeners predict a syllable to follow, it may also help them predict the tone type of that following syllable. If such a predictive effect on phonological form existed, input of a T3S modifier (e.g., *zhu2 sun3→2…* 'bamboo-shoot') would trigger a T3 head noun (e.g., *zhu2 sun3* 'bamboo-shoot' + r*u3 niu2* 'cow' → *zhu2 sun2 ru3 niu2*), while input of a surface T3 modifier (e.g., *zhu2 sun3…* 'bamboo-shoot') would trigger a non-T3 head noun (e.g., *zhu2 sun3* 'bamboo-shoot' + *hu2 die2* 'butterfly'→ *zhu2 sun3 hu2 die2*). However, no evidence was found that the tone type of the head was predictable based on the tone type of the modifier. These results provided an answer to the second critical research question: T3S does not help listeners to predict phonological form.

The results of Experiment 1 and Experiment 2 suggest that listeners can use T3S to predict lexical structure but not phonological form. This study therefore suggests that listeners treat the processing of lexical structure and phonological form as different types of processes: the former as structural computation and the latter as the retrieval of lexically stored information.

7.2 The value of using T3S to examine the effect of prosodic information in lexical processing

T3S was the critical phenomenon in this study for two reasons. First, T3S can be used to examine whether the tone type of an upcoming component is predictable or not (cf. the CAR, which cannot be used for this purpose). Second, T3S application does not entail a one-to-one association with the morphosyntactic or prosodic environment, which provides a good comparison with the predictive effect on lexical structure of the CAR, which does entail a strict one-to-one association.

The Mandarin T3S rule and the Japanese CAR differ in terms of their restrictions on the prosodic pattern of the upcoming component. The CAR applies to a compound word no matter what the original accentual pattern of its head is, and hence the accentual change derived from the CAR on a modifier cannot provide any information concerning the accentual pattern of the head noun. In contrast, when the T3S rule is applied, it is phonologically motivated: a T3S syllable must be followed by an underlying T3 syllable. Listeners may use the association to predict the tone type of the upcoming material when they encounter a T3S syllable as input (i.e., a modifier with a T3S syllable triggers the prediction of a head noun of the T3 type, while a modifier with a surface T3 syllable triggers the prediction of a head noun of a non-T3 type). This difference between the T3S rule and the CAR is why T3S offers a testing ground for the predictive effects on phonological form that the CAR cannot provide.

Furthermore, the CAR is obligatory and only applies to compound words. Hirose and Mazuka (2015, 2017) have shown that Japanese listeners use this one-to-one association as a cue to anticipate compound structure when they encounter input in which accentual change due to the CAR has occurred. In contrast to the CAR, the T3S rule is flexible and optional. These two uncertainties of T3S application conceivably cause processing ambiguities due to the resulting non-one-to-one association; listeners may avoid using such an association as a cue.

Concerning its flexibility, because the T3S rule is recursive it is difficult for listeners to know where to put a boundary, and what type of boundary it should be, after a T3S syllable. In the same sense, it is difficult for listeners to know whether a noun including a T3S syllable (e.g., *zhu2 sun3* →*2*… 'bamboo-shoot') is a modifier of a compound (e.g., *zhu2 sun3* 'bamboo-shoot' + *ru3 niu2* 'cow' → *zhu2 sun2 ru3 niu2*), or is a single noun that will be followed by a predicate (e.g., *zhu2 sun3* 'bamboo-shoot' + *hen3 xiang1* 'very fragrant' → *zhu2 sun2 hen3 xiang1*). This flexibility in T3S application may interfere with its usefulness for predicting lexical structure.

Second, the optionality of T3S application is observed in native uses. That is, an underlying T3 syllable in a T3S environment is allowed to remain unchanged (i.e., a T3S environment such as *zhu2 sun3* 'bamboo-shoot' + *ru3 niu2* 'cow' can be realized as output with an unchanged T3 syllable, *zhu2 sun3 ru3 niu2*, or as output with a T3S syllable, *zhu2 sun2 ru3 niu2*). In other words, if listeners are forced to choose between single (e.g., *zhu2 sun3* 'bamboo-shoot') and compound structure (e.g., *zhu2 sun3* 'bamboo-shoot' + *ru3 niu2* 'cow'), an input *zhu2 sun3* may not help listeners eliminate the possibility of compound structure. Therefore, the optionality of T3S application may interfere with its ability to predict lexical structure as well.

However, the results of Experiment 1 (Chapter 5) suggested that despite the T3S rule being flexible while the CAR is not, Mandarin listeners are capable of using T3S as a prosodic cue in predicting lexical structure, just as Japanese listeners can use the CAR.

7.3 Suggestions for further research

Many aspects of T3S deserve further study, for example, the association between the T3S and the types of structural boundary. As discussed in Section 3.2.3, T3S application is optional, but must be controlled by a set of factors. Previous studies have shown that an unchanged T3 syllable in a T3S environment may occur more readily at a foot, word, or phrase boundary. However, in a syllable sequence that can be interpreted with different types of boundary, it is unclear whether boundary type affects whether a T3S syllable or an unchanged T3 syllable occurs.

In Mandarin Chinese, a coordinate conjunction such as *he2* 'and' [和] is not necessary to express a parallel connecting relationship (e.g., [[*wo3*]$_{\text{NP}}$ [*mai3 le0*]$_{\text{VP}}$ [*yi1 fu2*]$_{\text{NP}}$ [*pi2 bao1*]$_{\text{NP}}$ [*xie2 zi0*]$_{\text{NP}}$]$_{\text{S}}$, I/bought/clothes/bags/shoes [我買了衣服皮包鞋子], 'I bought clothes, bags, and shoes'). For this reason, a phrase like *zhu2 sun3* 'bamboo-shoot'+ *ru3 niu2* 'cow' can have two different structural interpretations: as an N-N compound [[*zhu2 sun3*]$_{\text{N}}$ [*ru3 niu2*]$_{\text{N}}$]$_{\text{N}}$ where *zhu2 sun3* is the modifier of the head *ru3 niu2*; or as an NP-

NP coordination [[*zhu2 sun3*]$_{\text{NP}}$ (and) [*ru3 niu2*]$_{\text{NP}}$]$_{\text{NP}}$, in which it is read as a list of two independent single nouns, *zhu2 sun3* and *ru3 niu2*. In other words, *sun3* in the N-N compound [[*zhu2 sun3*]$_{\text{N}}$ [*ru3 niu2*]$_{\text{N}}$]$_{\text{N}}$ has a word boundary while *sun3* in the NP-NP coordinate phrase [[*zhu2 sun3*]$_{\text{NP}}$ (and) [*ru3 niu2*]$_{\text{NP}}$]$_{\text{NP}}$ has a phrase boundary.

A phrase has a larger structure than a word. Hence, the [*zhu2 sun3*]$_{\text{NP}}$ in the N-N coordination [[*zhu2 sun3*]$_{\text{NP}}$ (and) [*ru3 niu2*]$_{\text{NP}}$]$_{\text{NP}}$ has a larger structure than the [*zhu2 sun3*]$_{\text{N}}$ in the N-N compound [[*zhu2 sun3*]$_{\text{N}}$ [*ru3 niu2*]$_{\text{N}}$]$_{\text{N}}$. A further experiment could examine whether an input of T3S or unchanged T3 can be used to predictively distinguish the type of boundary (word boundary or phrase boundary). That is, if listeners tend to mark a larger boundary with unchanged T3, input of an unchanged T3 syllable triggers an NP-NP coordination (i.e., at a phrase boundary) while input of a T3S syllable triggers an N-N compound (i.e., at a word boundary). If both cases were confirmed, these results would indicate that listeners tend to relate an unchanged T3 syllable to a larger structure while relating a T3S syllable to a smaller structure when these structural interpretations can be derived from the same sequence. That is, the realization of a T3S syllable or an unchanged T3 syllable could then be taken to have a role in marking a boundary at different structural levels.

7.4. Conclusion

Tone 3 sandhi has a predictive effect on lexical structure, but not on phonological form. The flexibility and optionality of the T3S rule make it useful for testing predictive effects, because they result in a non-one-to-one association between a T3S syllable and its environment. The CAR, in contrast, cannot be used in this way because it does have a one-to-one association with its environment. Further studies should be conducted to explore how different T3S applications are associated with types of boundary. Future findings on the use of such associations in online predictive processing should contribute to our theoretical understanding of how T3S is generated in different structures.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Atkinson-King, K. (1973). Children's acquisition of phonological stress contrasts. *UCLA Working Papers in Phonetics, 25*.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Bahill, A. T., & McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research*, *23*(12), 1573-1583.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex?. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 50-68.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24-27.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, *48*, 62-69.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese* (Vol. 4). The Hugue: Mouton.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*(3), 201-220.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113-121.

Davis, M. H. (2000). *Lexical segmentation in spoken word recognition*. (Unpublished doctoral dissertation). Birkbeck College, University of London.

Dell, F. (2004). On unwarranted claims about stress and tone in Beijing Mandarin. *Cahiers de Linguistique Asie Orientale (Paris)*, *31*(1), 33-63.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121.

Duanmu, S. (2007). *The phonology of standard Chinese (2nd ed.)*. Oxford University Press.

Findlay, J. M., & Gilchrist, I. D. (2006). *Active vision: The psychology of looking and seeing* (No. 37) *アクティブ・ビジョン：眼球運動の心理・神経科学*. (H. Honda 本田仁視, Trans.). Kyoto: Kitaohji Press. (Original work published 2003)

Gart, J. J., Pettigrew, H. M., & Thomas, D. G. (1985). The effect of bias, variance estimation, skewness and kurtosis of the empirical logit on weighted least squares analyses. *Biometrika*, *72*(1), 179-190.

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from Japanese. *Cognition*, *136*, 350-358.

Hirose, Y., & Mazuka, R. (2017). Exploiting pitch accent information in compound processing: A comparison between adults and 6-to 7-year-old children. *Language Learning and Development*, *13*(4), 375-394.

Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics 55*, 23-42.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1-11.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541-573.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156.

Kubozono, H. (2008). Japanese accent. In S. Miyagawa and M. Saito (Eds.), *The Oxford Handbook of Japanese Linguistics* (pp. 165-191), Oxford University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), doi: 10.18637/jss.v082.i13.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549-565.

Misono, Y., Mazuka, R., Kondo, T., & Kiritani, S. (1997). Effects and limitations of prosodic and semantic biases on syntactic disambiguation. *Journal of Psycholinguistic Research*, *26*(2), 229-245.

Peng, S. H. (2000). Lexical versus ‘phonological’ representations of Mandarin sandhi tones. In Broe, M. B. and Pierrehumber. J. (Eds.). *Papers in laboratory phonology V: Acquisition and the lexicon* (pp.152-167). Cambridge University Press.

Shen, J. (1994). Běi jīng huà shǎng shēng lián xù de diào xíng zǔ hé hé jiē zòu xíng shì [Tonal patterns and rhythmic structure in successive third tones in the Beijing dialect]. *Zhong-guo Yuwen*, *4*, 274-281.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *The Journal of the Acoustical Society of America*, *133*(5), 3016-3029.

Shih, C. (1986). *The Prosodic Domain of Tone Sandhi in Chinese*. (Unpublished doctoral dissertation). University of California, SanDiego.

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. In Wang J. and Norval S. (Eds.), *Studies in Chinese Phonology* (pp. 81-123). Berlin: Mouton de Gruyter.

Speer, S., & Xu, L. (2008). Processing lexical tone in third-tone sandhi. *Laboratory Phonology*, *11*, 131-132.

Traxler, M., & Gernsbacher, M. A. (Eds.). (2011). *Handbook of psycholinguistics*. Elsevier.

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language*, *29*(2), 225-250.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61-83.

Yip, M. (2002). *Tone*. Cambridge University Press.

Yuan, J. and Chen, Y. (2014). 3rd tone sandhi standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, *42*(1), 218-236.

Zhang, J. (2014). Tones, tonal phonology, and tone sandhi. in Huang, C.-T. J., Li, Y.-H. A.& Simpson, A. (Eds.). *The Handbook of Chinese Linguistics* (pp. 443-464), Oxford.

Appendix A

Reading list used in production task (Chapter 4)

1	蘋果野豬	*ping2 guo3 + ye3 zhu1*	“apple wild-boar”
2	西瓜乳牛	*xi1 gua1 + ru3 niu2*	“watermelon cow”
3	蘿蔔	*luo2 bo1**[21]	“radish”
4	青椒老鷹	*qing1 jiao1 + lao3 ying1*	“green-pepper eagle”
5	青椒乳牛	*qing1 jiao1 + ru3 niu2*	“green-pepper cow”
6	鳳梨老鷹	*feng4 li2 + lao3 ying1*	“pineapple eagle”
7	蘋果蟒蛇	*ping2 guo3 + mang3 she2*	“apple python”
8	蘋果乳牛	*ping2 guo3 + ru3 niu2*	“apple cow”
9	鳳梨	*feng4 li2*	“pineapple”
10	南瓜乳牛	*nan2 gua1 + ru3 niu2*	“pumpkin cow”
11	香蕉蟒蛇	*xiang1 jiao1 + mang3 she2*	“banana python”
12	蕃薯浣熊	*fan1 shu3 + wan3* xiong2*	“sweet-potato raccoon”
13	青椒	*qing1 jiao1*	“green-pepper”
14	蓮藕浣熊	*lian2 ou3 + wan3* xiong2*	“lotus-root raccoon”
15	芒果蟒蛇	*mang2 guo3 + mang3 she2*	“mango python”
16	南瓜海豚	*nan2 gua1 + hai3 tun2*	“pumpkin dolphin”
17	竹筍	*zhu2 sun3*	“bamboo-shoot”
18	竹筍蟒蛇	*zhu2 sun3 + mang3 she2*	“bamboo-shoot python”
19	竹筍浣熊	*zhu2 sun3 + wan3* xiong2*	“bamboo-shoot raccoon”
20	蘿蔔海豚	*luo2 bo1* + hai3 tun2*	“radish dolphin”
21	西瓜浣熊	*xi1 gua1 + wan3* xiong2*	“watermelon raccoon”
22	蘋果	*ping2 guo3*	“apple”
23	芒果	*mang2 guo3*	“mango”
24	西瓜	*xi1 gua1*	“watermelon”

---

[21] *: The critical items (plant or animal) which are difference in pronunciation in Taiwan Mandarin and Beijing Mandarin:

Mandarin Character	translation	Taiwan Mandarin	Beijing Mandarin
浣熊	raccoon	*wan3 xiong2*	huan4 xiong2
企鵝	penguin	*qi4 e2*	*qi3 e2*
蘿蔔	radish	*luo2 bo1*	*luo2 bo0*

25	南瓜	*nan2 gua1*	“pumpkin”
26	鳳梨海豚	*feng4 li2 + hai3 tun2*	“pineapple dolphin”
27	鳳梨浣熊	*feng4 li2 + wan3* xiong2*	“pineapple raccoon”
28	玉米乳牛	*yu4 mi3 + ru3 niu2*	“corn cow”
29	玉米老鷹	*yu4 mi3 + lao3 ying1*	“corn eagle”
30	南瓜野豬	*nan2 gua1 + ye3 zhu1*	“pumpkin wild-boar”
31	西瓜蟒蛇	*xi1 gua1 + mang3 she2*	“watermelon python”
32	蓮藕蟒蛇	*lian2 ou3 + mang3 she2*	“lotus-root python”
33	蕃薯	*fan1 shu3*	“sweet-potato”
34	芒果海豚	*mang2 guo3 + hai3 tun2*	“mango dolphin”
35	青椒海豚	*qing1 jiao1 + hai3 tun2*	“green-pepper dolphin”
36	香蕉	*xiang1 jiao1*	“banana”
37	蓮藕海豚	*lian2 ou3 + hai3 tun2*	“lotus-root dolphin”
38	香蕉老鷹	*xiang1 jiao1+ lao3 ying1*	“banana eagle”
39	青椒浣熊	*qing1 jiao1 + wan3* xiong2*	“green-pepper raccoon”
40	鳳梨野豬	*feng4 li2 + ye3 zhu1*	“pineapple wild-boar”
41	蓮藕	*lian2 ou3*	“lotus-root”
42	西瓜野豬	*xi1 gua1 + ye3 zhu1*	“watermelon wild-boar”
43	香蕉浣熊	*xiang1 jiao1+ wan3* xiong2*	“banana raccoon”
44	蓮藕野豬	*lian2 ou3 + ye3 zhu1*	“lotus-root wild-boar”
45	竹筍海豚	*zhu2 sun3 + hai3 tun2*	“bamboo-shoot dolphin”
46	蕃薯乳牛	*fan1 shu3+ ru3 niu2*	“sweet-potato cow”
47	蕃薯海豚	*fan1 shu3+ hai3 tun2*	“sweet-potato dolphin”
48	玉米浣熊	*yu4 mi3 + wan3* xiong2*	“corn raccoon”
49	蘿蔔野豬	*luo2 bo1* + ye3 zhu1*	“radish wild-boar”
50	玉米野豬	*yu4 mi3 + ye3 zhu1*	“corn wild-boar”
51	蘋果老鷹	*ping2 guo3 + lao3 ying1*	“apple eagle”
52	芒果老鷹	*mang2 guo3 + lao3 ying1*	“mango eagle”
53	玉米	*yu4 mi3*	“corn”
54	蘿蔔老鷹	*luo2 bo1* + lao3 ying1*	“radish eagle”
55	南瓜蟒蛇	*nan2 gua1+ mang3 she2*	“pumpkin python”
56	蕃薯老鷹	*fan1 shu3 + lao3 ying1*	“sweet-potato eagle”
57	芒果野豬	*mang2 guo3 + ye3 zhu1*	“mango wild-boar”
58	香蕉乳牛	*xiang1 jiao1+ ru3 niu2*	“banana cow”
59	蘿蔔蟒蛇	*luo2 bo1* + mang3 she2*	“radish python”
60	竹筍乳牛	*zhu2 sun3 + ru3 niu2*	“bamboo-shoot cow”

Appendix B

Experiment items used in experiment 1 (Chapter 5)

(15a) T3/single

蘋果	*ping2 guo3*	“apple”
芒果	*mang2 guo3*	“mango”
蓮藕	*lian2 ou3*	“lotus-root”
竹筍	*zhu2 sun3*	“bamboo-shoot”
蕃薯	*fan1 shu3*	“sweet-potato”
玉米	*yu4 mi3*	“corn”

(15b) T3/compound

蘋果	*ping2 guo3*	“apple”
芒果	*mang2 guo3*	“mango”
蓮藕	*lian2 ou3*	“lotus-root”
竹筍	*zhu2 sun3*	“bamboo-shoot”
蕃薯	*fan1 shu3*	“sweet-potato”
玉米	*yu4 mi3*	“corn”

×

老鷹	*lao3 ying1*	“eagle”
野豬	*ye3 zhu1*	“wild-boar”
蟒蛇	*mang3 she2*	“python”
海豚	*hai3 tun2*	“dolphin”
浣熊	*wan3* xiong2*	“raccoon”
乳牛	*ru3 niu2*	“cow”

(15c) non-T3/single

香蕉	*xiang1 jiao1*	“banana”
鳳梨	*feng4 li2*	“pineapple”
西瓜	*xi1 gua1*	“watermelon”
青椒	*qing1 jiao1*	“green-pepper”
蘿蔔	*luo2 bo1**	“radish”
南瓜	*nan2 gua1*	“pumpkin”

(15d) non-T3/compound

香蕉	*xiang1 jiao1*	“banana”
鳳梨	*feng4 li2*	“pineapple”
西瓜	*xi1 gua1*	“watermelon”
青椒	*qing1 jiao1*	“green-pepper”
蘿蔔	*luo2 bo1**	“radish”
南瓜	*nan2 gua1*	“pumpkin”

×

老鷹	*lao3 ying1*	“eagle”
野豬	*ye3 zhu1*	“wild-boar”
蟒蛇	*mang3 she2*	“python”
海豚	*hai3 tun2*	“dolphin”
浣熊	*wan3* xiong2*	“raccoon”
乳牛	*ru3 niu2*	“cow”

Appendix C

Experiment items used in experiment 2 (Chapter 6)

(21a) T3S + T3

蘋果	_ping2 guo3_	“apple”		老鷹	_lao3 ying1_	“eagle”
芒果	_mang2 guo3_	“mango”		野豬	_ye3 zhu1_	“wild-boar”
蓮藕	_lian2 ou3_	“lotus-root”	×	蟒蛇	_mang3 she2_	“python”
竹筍	_zhu2 sun3_	“bamboo-shoot”		海豚	_hai3 tun2_	“dolphin”
蕃薯	_fan1 shu3_	“sweet-potato”		浣熊	_wan3* xiong2_	“raccoon”
玉米	_yu4 mi3_	“corn”		乳牛	_ru3 niu2_	“cow”

(21b) T3 + non-T3

蘋果	_ping2 guo3_	“apple”		蝴蝶	_hu2 die1_	“butterfly”
芒果	_mang2 guo3_	“mango”		企鵝	_qi4* e2_	“penguin”
蓮藕	_lian2 ou3_	“lotus-root”	×	烏龜	_wu1 gui1_	“tortoise”
竹筍	_zhu2 sun3_	“bamboo-shoot”		大象	_da4 xiang4_	“elephant”
蕃薯	_fan1 shu3_	“sweet-potato”		綿羊	_mian2 yang2_	“sheep”
玉米	_yu4 mi3_	“corn”		斑馬	_ban1 ma3_	“zebra”

(21c) non-T3 + T3

香蕉	_xiang1 jiao1_	“banana”		老鷹	_lao3 ying1_	“eagle”
鳳梨	_feng4 li2_	“pineapple”		野豬	_ye3 zhu1_	“wild-boar”
西瓜	_xi1 gua1_	“watermelon”	×	蟒蛇	_mang3 she2_	“python”
青椒	_qing1 jiao1_	“green-pepper”		海豚	_hai3 tun2_	“dolphin”
蘿蔔	_luo2 bo1*_	“radish”		浣熊	_wan3* xiong2_	“raccoon”
南瓜	_nan2 gua1_	“pumpkin”		乳牛	_ru3 niu2_	“cow”

(21d) non-T3 + non-T3

香蕉	_xiang1 jiao1_	“banana”		蝴蝶	_hu2 die1_	“butterfly”
鳳梨	_feng4 li2_	“pineapple”		企鵝	_qi4* e2_	“penguin”
西瓜	_xi1 gua1_	“watermelon”	×	烏龜	_wu1 gui1_	“tortoise”
青椒	_qing1 jiao1_	“green-pepper”		大象	_da4 xiang4_	“elephant”
蘿蔔	_luo2 bo1*_	“radish”		綿羊	_mian2 yang2_	“sheep”
南瓜	_nan2 gua1_	“pumpkin”		斑馬	_ban1 ma3_	“zebra”