UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

CHINESE COMPOUNDS: THE ROLE OF MORPHOSYNTACTIC STRUCTURE IN STRESS ASSIGNMENT IN SHANGHAI CHINESE AND TONE SANDHI IN MANDARIN CHINESE

Liu Haiyang

Dissertation supervised by Professor Marina Vigário, in Fulfillment of the Requirements for the Degree of Master of Linguistics

2022

Acknowledgement

“To be truly educated means to be in a position to inquire and to create on the basis of the resources available to you which you've come to appreciate and comprehend. To know where to look, to know how to formulate serious questions, to question a standard doctrine if that's appropriate, to find your own way, to shape the questions that are worth pursuing, and to develop the path to pursue them. That means knowing, understanding many things but also, much more important than what you have stored in your mind, to know where to look, how to look, how to question, how to challenge, how to proceed independently, to deal with the challenges that the world presents to you and that you develop in the course of your self-education and inquiry and investigations, in cooperation and solidarity with others.”
—Noam Chomsky

To complete this thesis work, special thanks go to Prof. Marina Vigário for her time, help and freedom given to me during the whole process.
Thanks to Prof. Sónia Frota, Prof. Telmo Móia, Prof. Gabriela Matos for their help and support in the courses of the master program and during the process of thesis.
Thanks to my family and friends for their support and help.
Thanks to all the infinite possibilities, updates and progress that have happened, are happening and will happen in our life.

Table of contents

Abstract

At the interface of morphosyntax and phonology, some phonological behaviors in Chinese languages are sensitive to word domain (stress assignment/stress resolution and tone sandhi). In this thesis, we focus on how morphosyntactic structures can contribute to some phonological behaviors that remain to be puzzles in the Chinese languages. Additionally, a highly-functional morphosyntax-based framework is shown to be realistic to construct a simplified and consistent model in domain construction of T3 tone sandhi in Chinese Mandarin, which has been considered challenging in the literature.

Following "Little x heads" theory (Marantz 1995; Marantz 2001) and syntactic incorporated compounding structures (Harley 2009), we use a syntactic multiple-root incorporated structure for Chinese compounding structures to account for the stress assignment and stress resolution (stress clash avoidance) in Shanghai Chinese with revised Phase Impenetrability for Phonology (rPIP) (Embick 2013). Meanwhile, a tentative Concatenation rule (Pak 2008; Chen 2018) after Linearization of Morphological words is proposed to account for the domain construction in T3 tone sandhi in Mandarin Chinese, which refers to specific morphosyntactic information (morphosyntactic locality characteristics and c-command relations). Different from the literature, we add the syntactic multiple-root incorporated structure of Chinese compounding structures into the algorithm of Concatenation rule. This is proved to be essential to successfully construct a unified framework of T3 tone sandhi in Mandarin Chinese both above and below the classical word domain, showing a noteworthy ability to deal with the exceptional situations in Chen (2009), e.g., syntactic words, phonological words and complex predicates. This project supports that morphosyntax-based analysis under syntactic word formation, e.g., Concatenation rules in Distributed Morphology, is a powerful weapon to reveal the processing logic of some controversial phonological rules vaguely floating between the classical lexical and postlexical rules in the literature, e.g., sandhi behaviours.

Under the current framework, differently from multimorphemic structures, the monomorphemic structures seem to be opaque in the application process of specific non-cyclic phonological rules. Such opaque monomorphemic structures can be postulated to be a product or outcome of certain phonological rules' processing economy and efficiency, instead of a true grammatical identity.

Key words: compounding structure; syntactic word formation; stress assignment; tone sandhi; Chinese languages

Resumo

A formação de palavras parece desempenhar um papel importante em comportamentos fonológicos em todas as línguas chinesas. Em Chen (2009), a formação de palavras surge como tendo forte influência na atribuição e resolução de acentos, em particular, no evitar de choques de acentos em chinês de Xangai e na ocorrência de sândi tonal envolvendo T3 em chinês mandarim.

a. O limite do domínio da palavra clássica parece ser problemático para o sandhi tonal que afeta sequências de T3 no chinês mandarim. As sequências de T3 no chinês mandarim têm o comportamento típico de um fenómeno de sandhi de natureza dissimilatória: um T3 na forma de citação é alterado para T2 se seguido por outra forma de citação com T3: T3—>T2/___T3 (Chen 2018). Em Chen (2009), a construção do domínio do sandhi tonal em chinês mandarim é dividida num processo em duas etapas—construção lexical e pós-lexical de MRU (unidade de ritmo mínima). Os domínios de sandhi tonal de T3 no chinês mandarim são construídos separadamente abaixo e acima do domínio de palavras clássicas, onde o domínio lexical de sandhi precisa ser construído estritamente antes do pós-lexical. Esse processo em duas etapas tem muitas exceções envolvendo o domínio da palavra clássica, que suscitam discussões excepcionais sobre palavra fonológica, palavra sintática e predicado complexo, que precisa ser visto como domínio lexical sob essa abordagem. (exemplos concretos sobre esse fenómeno encontram-se na Seção 2.4.2).

b. A distinção entre diferentes tipos de estruturas compostas, nomeadamente estruturas compostas formadas por "Substantivo-Substantivo" e "Substantivo-Adjective", parece ser essencial para explicar o fenómeno de choques de acentos assimétricos em chinês de Xangai. A atribuição de acentos em chinês de Xangai é geralmente reconhecida como tendo um padrão de distribuição de acentos à núcleo esquerda: "a proeminência à esquerda tem sido geralmente assumida para chinês de Xangai (Chen 2009)". Para a Chen (2009 : 307), seguindo Duanmu (1991, 1992a, 1993a), a atribuição de acentos em chinês de Xangai segue a seguinte regra (optámos aqui por manter a língua original):

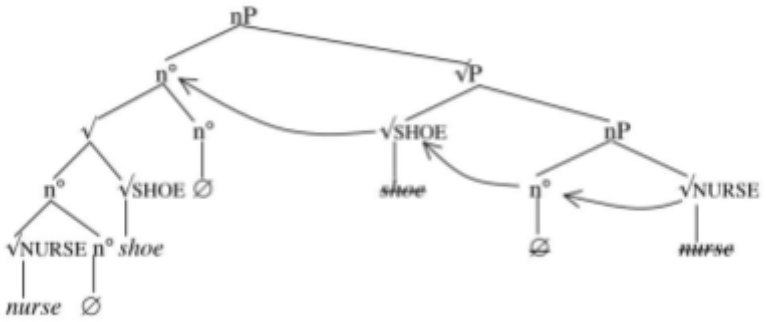
a. Morpheme level:
Line 0: trochee, left to right, ignore degenerate foot
Line 1: left-headed, unbounded stress
b. Word/compound level: assign cyclic left-headed stress

c. Phrase level: assign cyclic right-headed stress.
d. Stress Reduction: optionally delete line 1 stress
e. Clash Resolution: remove the stress column next to a higher column

Um choque de acentos surge quando duas sílabas consecutivas são acentuadas dentro do mesmo domínio fonológico (Angeliki 2018). Chen (2009) refere que a resolução de choques de acentos na atribuição de acentos em chinês de Xangai mostra um comportamento assimétrico em palavras de quatro morfemas—$[[[AB]_{N1}C_{N2}]D_{N3}]/[[A_{Adj1}[BC]_{N1}]D_{N2}]$. O choque de acentos com direcionalidade à esquerda é resolvido entre A_{Adj1} e $[BC]_{N1}$ na estrutura de quatro morfemas $[[A_{Adj1}[BC]_{N1}]D_{N2}]$, e o choque de acentos com direcionalidade à direita é evitado entre C_{N2} e D_{N3} na estrutura de quatro morfemas $[[AB]_{N1}C_{N2}]D_{N3}]$. (exemplos concretos sobre este fenómeno encontram-se na Seção 2.4.1)

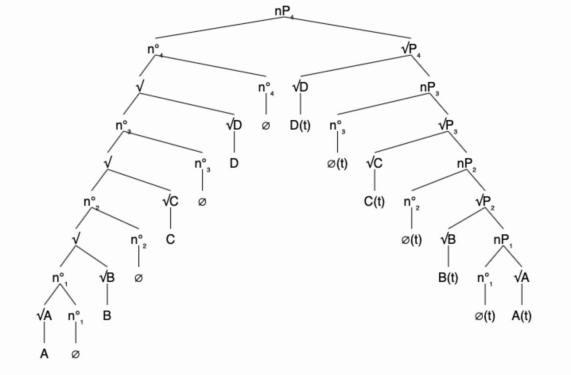
Para investigar o papel da estrutura morfossintática das estruturas compostas chinesas com os dados do sândi tonal envolvendo T3 e a atribuição de acentos em chinês de Xangai na interface de morfossintaxe e fonologia, usamos as ferramentas teóricas em Morfologia Distribuída: a. formação sintática de palavras e teoria de "*Little x heads*" (Marantz 1995; Marantz 2001). b. estruturas sintáticas de incorporação de compostos (Harley 2009), c. Impenetrabilidade de Fase para Fonologia revista (rPIP) (Embick 2013). d. regra de concatenação (Pak 2008).
Especificamente, adotando a estrutura de Morfologia Distribuída, proponho analisar estruturas compostas chinesas como estruturas sintáticas de incorporação de compostos de múltiplas raízes—em padrão Consecutivo ou Separado, adaptado de estruturas sintáticas de incorporação de compostos (Harley 2009) com formação sintática de palavras e teoria "*Little x heads*" (Marantz 1995; Marantz 2001).
A formação sintática de palavras propõe que o processo de formação de palavras é derivado de "Raiz" (e.g., √CAT) e morfemas abstratos (e.g., [Passado] ou [pl]) submetidos a operações sintáticas, como "*mover or merge*" (Embick e Noyer 2007; Embick 2010) e o morfema raiz incorpora nas suas núcleos funcionais definidoras de categorias cíclicas (v°, n°, a°) (Marantz 1995; Marantz 2001). Estruturas sintáticas de incorporação de compostos propõem que as palavras compostas clássicas são "estruturas incorporadas, onde substantivos não-núcleo se incorporam à raiz acategorial do substantivo principal, antes da sua própria incorporação em seu núcleo n° definidor de categoria" (Harley 2009), e.g., em "*nurse shoe*" em (2) abaixo, o substantivo não-núcleo *nurse* é incorporado em √*shoe*, antes da sua incorporação no n°.

A estrutura sintática de incorporação de compostos de duas raízes, retirada de Harley (2006: 17).



Eu adapto a abordagem de formação sintática de palavras—estruturas sintáticas de incorporação de compostos em Harley (2006) em uma estrutura de incorporação de múltiplas raízes, que repete estruturas sintáticas de incorporação de compostos em Harley (2006) em várias etapas – no padrão *Consecutive* or *Separate*, correspondente aos estruturas de quatro morfemas em chinês em compostos "Substantivos-Substantivos": $[[AB]_{N1}C_{N2}]D_{N3}$, ou em compostos "Adjectivos-Substantivos": $[A_{Adj1}[BC]_{N1}]D_{N2}$ /$A_{Adj1}[B_{Adj2}[CD]_{N1}]$, respectivamente. A estrutura de incorporação consecutiva de múltiplas raízes de $[[AB]_{N1}C_{N2}]D_{N3}$ é ilustrada em abaixo, onde o padrão consecutivo de movimentos de incorporação ocorre de √A —> √B —> √C —> √D:

Estrutura de incorporação consecutiva de múltiplas raízes de $[[AB]_{N1}C_{N2}]D_{N3}$:



Testamos estruturas sintáticas de incorporação de compostos com comportamentos fonológicos relevantes:

Sobre a construção do domínio de sândi tonal de T3 em chinês mandarim não cíclico, reconstruo uma regra de concatenação (Pak 2008; Chen 2018) referindo-se a informações

morfossintáticas específicas não locais (características de localidade morfossintática e relação c-comando). A principal diferença em relação ao que aparece na literatura em estudos sobre *Chaozhou Chinese* ou *Taiwan Southern Min* é que eu adiciono a estrutura com incorporação sintática de compostos ao algoritmo. Especificamente, o algoritmo usado em Chen (2018) aplica-se puramente no nível da frase, sem considerar o domínio clássico da palavra (por exemplo, a estrutura de compostos). A informação morfossintática específica que usamos é a relação de c-comando e características de localidade—relações entre o complemento e a núcleo ou posições de núcleos independentes. Quando adicionamos estruturas sintáticas de incorporação de compostos em questão, tais características entre os morfemas dentro do domínio lexical são adicionadas ao resultado do algoritmo integral da regra de concatenação para a construção do domínio de sândi, contribuindo para eliminar alguns resultados controversos da análise ao nível da frase pura. Esta proposta especial e importante de estruturas sintáticas de incorporação de compostos revela-se informação morfossintática-chave para unificar o algoritmo de regra de concatenação acima e abaixo do limite clássico da palavra, o que tem sido considerado problemático na literatura.

Sobre a regra cíclica de atribuição de acentos em chinês de Xangai, seguindo Marvin (2013), adaptei a regra cíclica de atribuição de acentos em chinês de Xangai para aplicar em "*Little x heads*" para alcançar uma ciclicidade estrutural com formação sintática de palavras (Marantz 1995; Marantz 2001) e a Impenetrabilidade de Fase para Fonologia revisada (rPIP) (Embick 2013). Seguindo Marvin (2013), na estrutura de palavras de combinação de "raiz e núcleo de categoria" sob Morfologia Distribuída, argumentamos que a regra cíclica de atribuição de acentos em chinês de Xangai se aplica em cada "xp" na estrutura incorporada de raiz múltipla proposta acima (*little "xps"—*$nP_1\,nP_2\,nP_3nP_4$). Com a Impenetrabilidade de Fase para Fonologia revista (rPIP) (Embick 2013), o choque de acentos evitado entre C_{N2} e D_{N3} na estrutura de quatro morfemas $[[AB]_{N1}C_{N2}]D_{N3}]$ pode ser potencialmente explicada de maneira estrutural, o que significa que as grelhas métricas (os acentos) "atribuídas de forma síncrona" nas primeiras fases de spell-out (i.e., a fase de spell-out de C) estando presentes, mas não alteráveis na última fase cíclica (i.e., a fase de spell-out de D). (veja-se a discussão detalhada na Seção 2.2.)

Nesta tese, os dados das línguas chinesas fornecem evidências para:

1). Os benefícios da formação sintática de palavras —estruturas sintáticas de incorporação de compostos:

a. Mostra-se que a estrutura de palavras compostas "Substantivos-Substantivos" (e.g. $[[[AB]_{N1}C_{N2}]D_{N3}]$) em chinês é construída por um padrão consecutivo de movimentos de incorporação, formando uma estrutura consecutiva de incorporação de compostos de múltiplas raízes. A estrutura de palavras compostas "Adjetivos-Substantivos" (e.g., $[[A_{Adj1}[BC]_{N1}]D_{N2}]/[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$) é construída por um padrão separado de movimentos de incorporação. Os resultados mostram que a estrutura "Adjetivos-Substantivos" precisa de seguir rigorosamente o movimento de incorporação de "Adjetivo" para "Substantivo". Isso sugere que a estrutura "Adjetivos-Substantivos" ainda pertence ao domínio clássico "Palavra", conforme a estrutura "Modificador-Substantivo" que possui as características de palavra tal como considerada na literatura.

b. Com estruturas sintáticas de incorporação de compostos, a construção do domínio de sândi tonal de T3 em chinês mandarim é potencialmente simplificada numa regra unificada de concatenação após a linearização de palavras morfológicas (Pak 2008), tanto abaixo quanto acima do domínio de palavra clássica, com o limite do domínio de palavra clássico deixando de ser problemático na literatura, como era o tratamento envolvendo MRU (unidade de ritmo mínima) sob a Teoria da Otimalidade em Chen (2009).

c. Com os efeitos das fases de *spell-out* cíclicas, a regra cíclica de atribuição de acentos em chinês de Xangai é bem derivada nas estruturas sintáticas de incorporação de compostos, fornecendo uma explicação possível para o conflito entre o choque de acentos imperativo e o choque de acentos evitado, descrito na literatura.

2). Os benefícios da estrutura de construção do domínio de sândi tonal de T3 no chinês mandarim como regra de concatenação:

a. No quadro da regra de concatenação após a linearização de palavras morfológicas, a construção do domínio de sândi tonal de T3 em chinês mandarim mostra melhor explicação às situações excepcionais em abordagens alternativas e melhor correção na previsão nos limites do domínio de sândi tonal de T3.

b. O domínio de sândi tonal de T3 é determinado por uma regra de concatenação de palavras morfológicas referindo-se às informações morfossintáticas específicas não locais (características de localidade morfossintática e c-comando). Isso mostra que algumas regras fonológicas têm acesso à informação sintática não local, dando suporte à afirmação de que

informações sintáticas específicas devem ser transparentes para a fonologia na interface sintaxe-fonologia.

c. Sob o quadro atual com regras fonológicas específicas como a construção do domínio de sândi tonal de T3 em chinês mandarim como regra de concatenação após linearização de palavras morfológicas, diferentemente das estruturas multimorfémicas, as estruturas monomorfémicas acabam por ser os espaços opaços para tais regras de concatenação. Isso pode sugerir que ainda temos um conceito de "palavra" estruturalmente definido para tais estruturas monomorfémicas na atual teoria sintática de formação de palavras, que não é uma verdadeira identidade gramatical, mas um produto ou resultado da economia e eficiência de processamento de certas regras fonológicas.

Palavras-chave: estrutura de composição; formação sintática de palavras; atribuição de acentos; sandhi tonal; línguas chinesas

1. Introduction

Word formation has been considered to play an important role in many phonological behaviors across Chinese languages. In Chen (2009), word formation is reported to have strong influence, e.g., in stress assignment and stress resolution (stress clash avoidance) in Shanghai Chinese and in T3 tone sandhi in Mandarin Chinese.

a. The classical word domain boundary seems to be problematic in T3 tone sandhi in Mandarin Chinese.

T3 tone sandhi in Mandarin Chinese is a typical sandhi behavior: citation form T3 is changed to sandhi form T2 if followed by another citation form T3: T3—> T2/___T3. In Chen (2009), T3 tone sandhi domain construction in Mandarin Chinese is divided into a two-pass process—lexical and postlexical MRU[1](minimal rhythm unit) construction. T3 tone sandhi domains are constructed separately below and above classical word domain, where the lexical sandhi domain needs to be constructed strictly earlier than the postlexical one. This two-pass process also needs abundant exceptions focused on classical word domain boundaries, including exceptional discussions about Phonological word, Syntactic word and complex predicate that needs to be viewed as lexical domain under this approach. (concrete examples about this phenomenon can be found in Section 2.4.2)

b. Distinguishing between different types of compounding structures, namely Noun-Noun and Adj-Noun compounding structures, seem to be essential to account for the asymmetrical stress clash phenomenon in Shanghai Chinese.

Stress assignment in Shanghai Chinese is generally acknowledged to have left-headed stress pattern: “left prominence has been commonly assumed for Shanghai (Chen 2009)”. The general stress assignment rule in Shanghai Chinese is taken from Chen (2009):

(1) Shanghai Chinese Stress Assignment by Duanmu (1991, 1992a, 1993a), taken from Chen (2009 : 307)

a. Morpheme level:
Line 0: trochee, left to right, ignore degenerate foot
Line 1: left-headed, unbounded stress

[1] MRU (minimal rhythm unit)—the sandhi domain in Mandarin Chinese determined by Chen (2009)

b. Word/compound level: assign cyclic left-headed stress
c. Phrase level: assign cyclic right-headed stress.
d. Stress Reduction: optionally delete line 1 stress
e. Clash Resolution: remove the stress column next to a higher column

A stress clash arises when two consecutive syllables are stressed within the same phonological domain (e.g., phonological phrase) (Angeliki 2018). Chen (2009) reported that stress clash resolution in Shanghai Chinese Stress Assignment shows directional asymmetrical behavior in four-morpheme words—$[[[AB]_{N1}C_{N2}]D_{N3}]$[2]/$[[A_{Adj1}[BC]_{N1}]D_{N2}]$. The left directional stress clash is resolved between A_{Adj1} and $[BC]_{N1}$ in four-morpheme structure $[[A_{Adj1}[BC]_{N1}]D_{N2}]$, and the right directional stress clash is avoided between C_{N2} and D_{N3} in four-morpheme structure $[[AB]_{N1}C_{N2}]D_{N3}]$. (concrete examples about this phenomenon are given in Section 2.4.1)

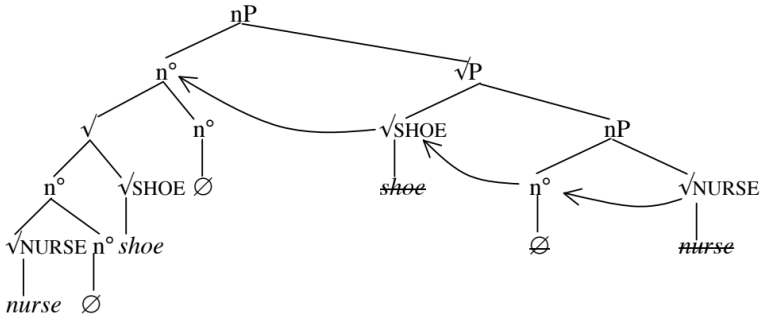
As we can see from the above, the currently available approaches to the phonological phenomena that are influenced by word domain or word structure can be systematically investigated in a deep way: **1).** Non-cyclic Tone sandhi in Mandarin Chinese: **a.** Dealing with this kind of phonological rules with both morphological and syntactic sensitivity in two separate application processes in lexical and post lexical levels can be a compromise. **b.** If we follow Chen (2009) to assume tone sandhi in Mandarin Chinese is a non-cyclic phonological rule, T3 tone sandhi rule is required to apply multiple times at lexical and postlexical level according to some requirements for the domain constructions, e.g., multiple domain construction in the cyclicity of compounding structures in the lexicon **c.** The domain construction process is more demanding to be practical when involving more exceptional cases (i.e., complex predicate — the structure of "verb+resultative verb" needs to be viewed as a lexical domain in the framework of Chen 2009). **2).** Cyclic Stress assignment in Shanghai Chinese: **a.** The asymmetric clash resolution due to the structure of compound word (cyclicity) seems not conventional in the context of Phonology. **b.** I agree with Chen (2009) that asymmetric clash resolution is "at best a merely inductive summary of the observed facts, at worst a desperate attempt to salvage the stress-based account".

[2] Following Chen (2009), we use bracketings to show the cyclicity in the classical lexicon. An example of $[[[AB]_{N1}C_{N2}]D_{N3}]$: $[[[\text{dou-fu}]_{N1}\text{gan}_{N2}]\text{si}_{N3}]$—$[[[\text{tofu}]_{N1}\text{dried-food}_{N2}]\text{slice}_{N3}]$— it shows in the classical lexicon, the $[\text{tofu}]_{N1}$ combines with $[\text{dried-food}_{N2}]$ firstly, then the $[[\text{tofu}]_{N1}\text{dried-food}_{N2}]$ combines with $[\text{slice}_{N3}]$ into $[[[\text{tofu}]_{N1}\text{dried-food}_{N2}]\text{slice}_{N3}]$. We use the bracketings to show the proximity between morphemes.

To investigate the role of morphosyntactic structure of Chinese compounding structures with the data of T3 tone sandhi in Mandarin Chinese and the stress assignment in Shanghai Chinese in the interface of morphosyntax and phonology, we are going to use the theoretical tools under Distributed Morphology: **a.** syntactic word formation and "Little x heads" theory (Marantz 1995; Marantz 2001). **b.** syntactic incorporated compounding structures (Harley 2009), **c.** revised Phase Impenetrability for Phonology (rPIP) (Embick 2013). **d.** Concatenation/Chaining phonological rules (Pak 2008)

Specifically, adopting the Distributed Morphology framework, I propose to analyze Chinese compounding structures under syntactic multiple-root incorporated structure—in Consecutive or Separate pattern, adapted from syntactic incorporated compounding structures (Harley 2009). Syntactic incorporated compounding structure argues that classical compound words are "incorporated structures, where non-head nouns incorporate into the acategorial root of the head noun, prior to its own incorporation into its category-defining n° head" (Harley 2009), e.g., In "nurse shoes" in (2) below, *nurse* is incorporated into √*shoes*, prior to its incorporation into n°.

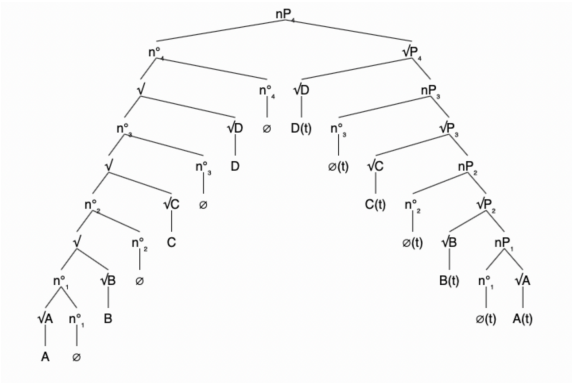
(2) Two-root incorporated compounding structure, taken from Harley (2006 : 17).



I adapt the syntactic word formation approach—incorporated compound word structure in Harley (2006) into a multiple-root incorporated structure, which repeats the incorporated compounding structure in Harley (2006) in multiple steps—in Consecutive or Separate pattern, corresponding to the four-morpheme[3] compounding structures in Chinese in pure Noun-Noun compounds: $[[AB]_{N1}C_{N2}]D_{N3}$ or Adj-Noun compounds: $[A_{Adj1}[BC]_{N1}]D_{N2}$ / $A_{Adj1}[B_{Adj2}[CD]_{N1}]$, respectively. The Consecutive multiple-root incorporated structure of $[[AB]_{N1}C_{N2}]D_{N3}$ is illustrated in (3) below, where the consecutive pattern of incorporation movements take place from √A —> √B —> √C —> √D:

[3] Following Chen(2009), we refer to the ambiguous "word" or syllable components in the compounding structures in Chinese languages as morphemes (three-morpheme/four-morpheme compounding structures etc.,).

(3) Consecutive multiple-root incorporated structure of $[[AB]_{N1}C_{N2}]D_{N3}$



We will test this syntactic compounding structure with the data from relevant phonological phenomena below:

About non-cyclic T3 Mandarin Chinese Tone sandhi domain construction, I reconstruct a Concatenation rule (Pak 2008; Chen 2018) referring to specific morphosyntactic information (morphosyntactic locality characteristics and c-command relation). The core difference from the literature with relevant studies about Chaozhou Chinese or Taiwan Southern Min is that I add the incorporated compounding structure into the algorithm. To be specific—the algorithm used in Chen (2018) is purely at the sentence level without consideration of classical word domain (e.g., the compounding structure). The specific morphosyntactic information we use is c-command relation and locality characteristics—head-complement relations or independent head positions. When we add the incorporated compounding structure into play, such characteristics between the morphemes inside the lexical domain are added into the integral algorithm result of the concatenation rule for sandhi domain construction, which makes it the key factor to rule out some controversial results from analysis at the pure sentence level. The syntactic compounding structure is shown to be the key morphosyntactic information to unify the algorithm of concatenation rule above and below the classical word boundary, which has been considered to be problematic in the literature.

About cyclic Shanghai Chinese Stress Assignment, following Marvin (2013), I adapted the cyclic Shanghai Chinese Stress assignment rule to apply in little "xps" to achieve a structural cyclicity with syntactic word formation (Marantz 1995; Marantz 2001) and revised Phase Impenetrability for Phonology (rPIP) (Embick 2013). Following Marvin (2013), in the standard "root and category-head" combination word structure under Distributed Morphology, we argue that cyclic stress assignment rule in Shanghai Mandarin applies in

every “xp” in the multiple-root incorporated structure in (3) above (little “*xps” nP_1 nP_2 nP_3 nP_4*). With revised Phase Impenetrability for Phonology (rPIP) (Embick 2013), the avoided right directional stress clash between C_{N2} and D_{N3} in four-morpheme structure $[[AB]_{N1}C_{N2}]D_{N3}]$ can be potentially accounted for in a structural way, which means the “assigned synchronously” metrical grids/stresses in the former spell-out phases (i.e., the spell out phase of C) being present but not changeable in the latter cyclic phase (i.e., the spell out phase of D). (see detailed discussion in Section 2.2)

We use the data about phonological behaviors (tone sandhi and stress assignment) in Chinese languages from Chen (2009), which is a classical work in Chinese Phonology. Chen (2009) has abundant and diversified examples to provide very detailed discussion, observation and analysis about phonological behaviors focused on tone sandhi behaviors across Chinese languages.

This thesis is arranged in the following sections: In section 2, some basic information from the theoretical background is given. In section 3, we propose a multiple-root incorporated structure of four-morpheme Chinese compounds adapted from Harley (2006). In section 4, we check each type of multiple-root incorporated structure in the Shanghai Stress assignment rule. In section 5, we propose a new framework of T3 tone sandhi domain construction as Concatenation rule in Mandarin Chinese, and formulate relevant algorithms with morphosyntactic information—morphosyntactic locality characteristics and c-command relations. In section 6, the multiple-root incorporated structure in section 3 is tested in the framework of T3 tone sandhi domain construction as Concatenation rule in Mandarin Chinese in section 5. In section 7, we summarize the general discussion.

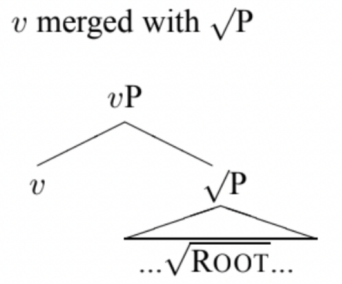
2.Theoretical background

2.1. From Distributed Morphology to syntactic incorporated structure of compounding structures

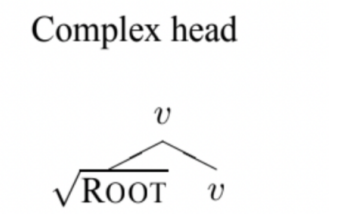
Syntactic-all-the-way-down morphology theory, e.g., Distributed morphology, argues that word formation process is derived from roots (e.g., √CAT) and abstract morphemes (e.g., [Past] or [pl]) undergoing syntactic operations, such as move or merge (Embick and Noyer 2007; Embick 2010). Meanwhile, roots (e.g., √CAT) are categorized by functional heads—“Little x heads” (v°, n°, etc. Marantz 1995; Marantz 2001). In Marantz (2001), the

functional category heads n°,v°,a° under word domain are determined to be cyclic heads, with respective complements being cyclically spelled-out. Embick (2010) provides concrete examples for this "root + category head" structure for "verb": "heads of this type categorize the elements that they attach to. So, for example, a head v which is merged syntactically to a √P headed by a category-neutral √ROOT creates a vP (4); when the Root and the v head are combined into a single complex head as shown in (5), the result is a "verb":

(4) Sample "root + category head" structure, taken from Embick (2010 : 9).

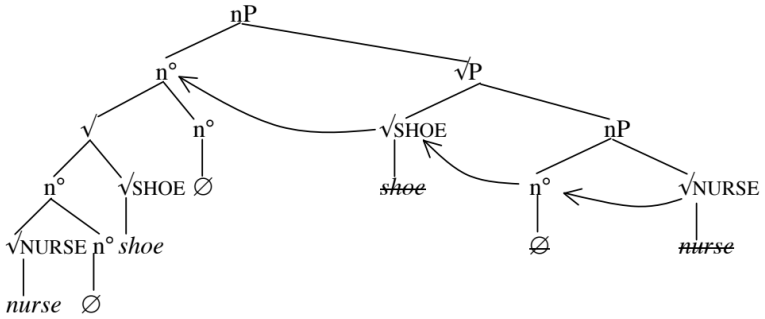


(5) "Verb" structure, taken from Embick (2010 : 9).



Under Distributed Morphology, incorporated compounding structure is a syntactic word formation process proposed for English syntactic and root (primary) compounding structures in Harley (2006). "Compounds are incorporated structures, where non-head nouns incorporate into the acategorial root of the head noun, prior to its own incorporation into its category-defining n° head" Harley (2006), e.g. In "nurse shoes", *nurse* is incorporated into √*shoes*, prior to its incorporation into n°. For convenience, we repeat (2) below:

(2) Two-root incorporated compounding structure, taken from Harley (2006 : 17).



2.2. Some phonological behaviors discussed in the literature related to syntactic word formation

Some phonological behaviors are discussed in the syntactic word formation proposal.
In Scott Jackson and Jeffrey Punske (2013), phrasal adjunct structure and complement (incorporation) structure are distinguished to account for initial-stress and final-stress between “black bírd” and “bláckbird”. They argue that the incorporated compound structure of “bláckbird” can account for the initial-stress pattern, adopting the sentencial prominence formulation with highest phase condition in Kratzer and Selkirk (2007).
In Marvin (2013), a general process about how the “root and category-head” combination word structure (e.g., [√govern *v*] ment *n*] al *a*] ese *n*]) interacting with the word stress assignment in English is given. He shows that the English stress shift behavior in word stress (e.g., góven-góvenment-góvenméntal-góvenméntalése) can provide evidence for syntactic word formation structure. The metrical grids are “assigned synchronously” under effects of cyclic spell-out phases, with former spell-out metrical grids/stresses being present but not changeable in the latter cyclic phase.

Halle (1998) proposes that “the English stress system is constituted by the Main Stress Rule supplemented by two Edge-Marking Rules. The Main Stress Rule has two parts. A binary foot is constructed at the end of a string whose last asterisk projects a light root. A unary foot is built if the last syllable is heavy or there are not enough syllables in the word to construct a binary foot”. Marvin (2013)’s adaptation of English cyclic stress assignment rule is shown below in (6). He argues that the English cyclic stress assignment rules, such as the Main Stress Rule and Edge-Marking Rules, are placed in each “xP” of syntactic word structure to naturally achieve the Cyclicity. The specific Prosodic word domain rules, e.g., Vowel

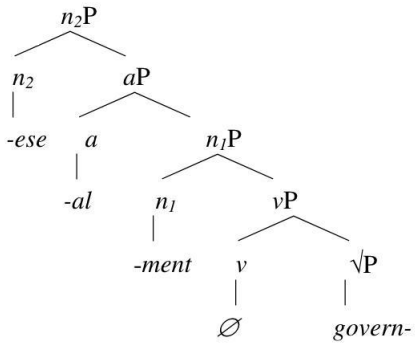
reduction rule, will be carried out only after the Prosodic word is constructed. In this way, the Cyclicity of the phonological rule application is potentially to be explained structurally.

(6) Marvin (2013)'s adaptation of English cyclic stress assignment rule, taken from Marvin (2013 : 8)

a. Words are composed of little *x*Ps, Marantz (2001);
b. MSR and EMR apply at every *x*P if triggered by a diacritic marking on *x* (i.e. by cyclic affixes); they also apply at the last *x*P if not triggered before;
c. Vowel Reduction Rule takes place at the level of 'prosodic word';
d. A 'phase analysis' (phase spell-out and PIC) as in (3) applies to stress in connection with vowel reduction.

We illustrate this process in (7):

(7) Syntactic structure for "góvernméntalése", taken from Marvin (2013 : 8)



1. At vP and n_1P, the Main Stress Rule and Edge-Marking Rules are applied, and "góvern" is spelled out.
2. At aP, the Main Stress Rule and Edge-Marking Rules are applied again, and n_1P—"góvernmént" is spelled out.
3. At n_2P, the Main Stress Rule and Edge-Marking Rules are applied again, and aP—"góvernméntal" is spelled out.
4. At a higher phrase, the Main Stress Rule, Edge-Marking Rules and Vowel Reduction Rule are applied again, and n_2P—"góvernméntalése" is spelled out.

In this way, the stress shift in English words is explained structurally. The metrical grids/stresses assigned by the English word stress rules applied in the former phases are not changeable in the latter phases. The English word stress rules applied in the latter phases keep

assigning new metrical grids and producing new word stresses to new syllables—leading to the phenomenon of stress shift in English words with multiple affixes.

This process in "góven-góvenment-góvenméntal- góvenméntalése" provides necessary information for us to propose the adaptation of the Shanghai Chinese stress assignment in the syntactic word structure: **a**. the metrical grids being assigned synchronously when every morpheme appears cyclically, during the construction of words. **b**. The metrical grids and the stresses determined by higher metrical grids structure spelled out in the former phases, cannot be altered in the latter phase of higher phase. **c**. The metrical grids in **b**. can be seen in the higher phase when cyclic rules apply again, and be treated as a part of structure in which cyclic rules apply again, but are not changeable in the latter phases. As we can understand , these metrical grids assigned in the former phases, contribute to the stress assignment in the later phases by just standing there during cyclic rules application in the higher phase.

2.3. Phase theory in syntactic word formation/Phase Impenetrability for Phonology

In Chomsky (2001), Phase Theory determines that vP and CP are phrases with cyclic domains, the complements of which are cyclically spelled out. In Embick (2010 and 2013), phase theory is extended into word domain with new discussion.

What is different from Phase Theory in the classical syntax is that the √root in deepest position will not be spelled out alone in the syntactic word formation structure. The cyclic spell-out domains in the syntactic word structure determined in Embick (2010) are shown in (8) below: The first cyclic "x" with √root as its complement will not choose to spell out its complement—the √root itself alone. √root will only be spelled out when merging with the second cyclic head "y". The cyclic "y" triggers the spell-out of its complement—√root, the cyclic "x" head, and materials between cyclic "x" and cyclic "y" (except the head "y" itself).

(8) The cyclic spell-out domains in the syntactic word structure, taken from Embick (2010 : 39)

a. $[[[[[\sqrt{\text{ROOT}}\ x]\ W]\ Z]}_{y}\ y]$

Cyclic y triggers spell out of cyclic domains in its complement. The head x undergoes VI, as do the edge+ heads W and Z. The Root is processed phonologically.

b. $[[[[[\sqrt{\text{ROOT}}\ \underbrace{x^*]\ W^*]\ Z^*]\ y]}_{z}\ \ldots\ z]$

Merge of higher cyclic z triggers spell out of cyclic domains in its complement. The head y defines a cyclic domain, and is subjected to VI (along with any edge+ heads it might have). The heads marked with $^*-x, W, Z-$ are present when y undergoes VI, but have undergone VI in the earlier cycle.

What is similar to Phase Theory in the classical syntax is that the complement of the former cyclic phase is "inactive" in the latter cyclic phase. In Revised **Phase Impenetrability for Phonology** (rPIP) in Embick (2013): "Material that is phase-cyclically inactive", **a.** "has a visible phonological representation, but cannot be identified as a particular morpheme", e.g. metrical grids assigned in the former cyclic phases are visible to the latter phases and continue to contribute to the whole metrical structure in the latter phases; **b.**"can only be seen and altered by non-cyclic or phrasal phonological rules", e.g. Vowel and consonant harmony and flapping, "but not by cyclic phonological rules", e.g. Stress assignment rule in English. "Tonal interactions, sandhi effects, etc." are mentioned particularly to be capable of breaking through the cyclic spell-out phases in word domain. (Embick 2013)

2.4. Brief information about cyclic stress assignment rule in Shanghai Chinese and non-cyclic tone sandhi rule in Mandarin Chinese

In **Phase Impenetrability for Phonology** in Embick (2013), cyclic and non-cyclic phonological rules show different levels of sensitivity to spell-out phases inside the word domain, cyclic stress assignment rule and non-cyclic tone sandhi rule are primarily chosen to check the Chinese compounding structure data in the syntactic word structure.

2.4.1. Cyclic stress assignment rule in Shanghai Chinese

Stress assignment in Shanghai Chinese is generally acknowledged to have left-headed stress pattern: "left prominence has been commonly assumed for Shanghai (Chen 2009)". The general stress assignment rule in Shanghai Chinese in (1) is repeated below in the interest of readability, and particularly the (e) stress clash resolution rule determines to remove the stress

column next to a higher column. In (10), a concrete example about stress clash resolution in Shanghai Chinese is taken from Chen (2009).

(1) Shanghai Chinese Stress Assignment by Duanmu (1991, 1992a, 1993a), taken from Chen (2009 : 307)

a. Morpheme level:
Line 0: trochee, left to right, ignore degenerate foot
Line 1: left-headed, unbounded stress
b. Word/compound level: assign cyclic left-headed stress
c. Phrase level: assign cyclic right-headed stress.
d. Stress Reduction: optionally delete line 1 stress
e. Clash Resolution: remove the stress column next to a higher column

(9) Asymmetrical stress clash resolution between $[[[AB]_{N1}C_{N2}]D_{N3}]$ and $[[A_{Adj1}[BC]_{N1}]D_{N2}]$, taken from Chen (2009 : 334)

Type **A**

Compound level: x
Word level: (x)(x)
Line 1: (x .)(x)(x)
Line 0: dou-fu-gan-si
"Shredded Dried Tofu"

Type **B**

Compound level: x
Word level: (x)(x)
Line 1: (x)(x .)(x)
Line 0: hu-luo-bo-si
"Shredded Carrot"

Compound level: x
Word level: (x)(x)
Line 1: (x . .)(x)
Line 0: hu-luo-bo-si
"Shredded Carrot

Due to different compounding structure structures, stress clash resolution shows asymmetrical performance in left and right directional stress clash in four-morpheme compounding structures—$[[[AB]_{N1}C_{N2}]D_{N3}]/[[A_{Adj1}[BC]_{N1}]D_{N2}]$. In (10), the (e) **left-directional** clash resolution rule in (9) is applied between "hu (x)" and "luo-bo (x .)" in type B word structure—$[[A_{Adj1}[BC]_{N1}]D_{N2}]$. The line 1 (foot-level) stress column is removed next to the next higher level, producing "hu-luo-bo-si (x . .)(x)". On the contrary, the (e) **right-directional** clash resolution rule in (9) is avoided between "gan (x)" and "si (x)" in type A word structure—$[[[AB]_{N1}C_{N2}]D_{N3}]$. The line 1 (foot-level) stress column remained to be "dou-fu-gan-si (x .)**(x)(x)**" with the stress clash avoided.

The structure of classical word structure (cyclicity) is controversial in cyclic Stress assignment in Shanghai Chinese: **a.** The asymmetric clash resolution seems not conventional in the context of Phonology. **b.** I agree with Chen (2009) that asymmetric clash resolution is "at best a merely inductive summary of the observed facts, at worst a desperate attempt to salvage the stress-based account".

2.4.2. Non-cyclic tone sandhi rule in Mandarin Chinese

T3 tone sandhi in Mandarin Chinese is a typical sandhi phenomenon: citation form T3 is changed to sandhi form T2 if followed by another T3: T3—> T2 / ___T3. (Chen 2018). In Chen (2009), T3 tone sandhi rule in Mandarin Chinese is divided into two-pass lexical and postlexical application process, including lexical and postlexical MRU (minimal rhythm unit) construction, which is the sandhi domain in Mandarin Chinese determined by Chen (2009). He shows that the classical word domain boundary seems to be problematic in the application of T3 tone sandhi rule in Mandarin Chinese: the MRU domain construction rule of T3 tone sandhi is applied multiple times—below and above classical word domain. For convenience, we take an example from Chen (2009) in (10) below, in [xiang [xie [xiaoshuo]]] ("plan to write a novel"): Lexical MRU (minimal rhythm unit) is firstly constructed by [xiaoshuo] ("novel"), where the lexical T3 tone sandhi applies (not applicable in this case). Postlexical/phrasal MRU (minimal rhythm unit) is secondly constructed by [xiang [xie [xiaoshuo]]] ("plan to write a novel"), where the obligatory postlexical/phrasal T3 tone sandhi applies between "xiang" ("plan") and "xie" ("write") and the optional cross-MRU (minimal rhythm unit) T3 tone sandhi applies between "xie" ("write") and "xiao-shuo" ("novel").

(10) Example of two-pass lexical and postlexical MRU (minimal rhythm unit) construction, taken from Chen (2009 : 386)

xiang # [*xie* # *xiao-shuo*]				"plan to write a novel"	
3	3	3	1	base tone	
		(3	1)	Lexical MRU, TS not applicable	
(s	3)	(3	1)	Phrasal MRU, TS	i = ok
(s	s)	(3	1)	Cross-MRU TS optional	j = ok

Chen (2009) has a very comprehensive display of data showing morphological and syntactic sensitivity. If we look at the dispute between lexical and phrasal, to our understanding, the two-pass sandhi rule application process seems to be a **compromise** to treat the

morphological and syntactic sensitivity of T3 tone sandhi in Mandarin Chinese: **a.** At lexical level, classical cyclicity has a decisive influence in determining tone sandhi domain boundaries in T3 tone sandhi in Mandarin Chinese. **b.** At the post-lexical level, the rankings of constraints in the context of Optimality Theory include more fatores, such as syllable quantity (e.g., The MRU is at least disyllabic or The MRU is at most disyllabic) and syntactic information (e.g., congruence—Group X forms an MRU with its closest morphosyntactic mate) etc., **c.** There are abundant exceptions. These points will be the focus of our current analysis.

2.4.3. Non-cyclic tone sandhi behaviours in Chinese languages in more frameworks—Precompiled Phonology/Distributed Morphology

More approaches in the literature have been used to study the abundant characteristics of tone sandhi behaviours across Chinese languages. The allomorphy-related frameworks provide a new perspective to look into the syntactic and morphological sensitivity of different tone sandhi behaviours. Hayes (1990) regards tone sandhi in Xiamen Chinese as a Precompiled Rule stored in the lexicon, which is an approach to explain that lexical phrasal rule share the characteristics of **both** the lexical rules, e.g. sensitivity to morphological information, **and** of the phrasal rules, e.g. sensitivity to syntactic information. He argues that the corresponding allomorphs of tone sandhi are generated lexically (in classical lexicon) in the relevant syntactic contexts but selected for insertion into the relevant phonological instantiation in the postlexical component. Hayes (1990) argues that in Xiamen Chinese, "where the relevant allomorphs are inserted at the right edges of non-adjunct maximal projections". Allomorphy-related framework can provide an alternative way to explain why the tone sandhi behaviours share the characteristics of both lexical phonological rule, e.g. morphological sensitivity (classical cyclicity)/existence of rich exceptions, and phrasal phonological rule, e.g. syntactic sensitivity (head or non-head position and syntactic word)/pause effects (IP bound).

More recently, allomorphy theory is combined with syntactic morphology, e.g. Distributed morphology. In the context of Distributed Morphology, the traditional lexicon was replaced by syntactic operation extended to the inside of word domain. Allomorphy is no longer assumed to be stored in the lexicon in the Precompiled Phonology and subsequently be inserted postlexically. However, it is directly determined by the adjacent morpheme (syllable). Pak (2008) argues that tone sandhi in Xiamen Chinese follows the allomorphy

selection as: “A. Insert allomorph X at the right edge of a clausal adjunct, if the following word is a vowel-initial adjective. B. Insert allomorph Y at the left edge of a parenthetical XP, if the word preceding the parenthetical is a bimoraic noun.” Pak (2008) also proposed for Chaozhou Chinese tone sandhi: “Chaozhou tone sandhi applies strictly from right to left across a string of words within a given domain, regardless of the syntactic bracketing.” “If a rule applies strictly right-to-left or left-to-right in a way that requires reference to a string of linearized words, it must apply after Chaining.” Importantly, in Pak (2008), two types of concatenation processes of Morphological words are introduced: 1. Head-left Concatenation: identifies pairs of Morphological words X, Y where (i) X is left-adjacent to Y, and (ii) X c-commands Y. 2. Phrase-left Concatenation: identifies pairs of M-words[4] X, Y where (i) X is left-adjacent to Y, and (ii) X does not c-command Y. This will be the basic function of the algorithm in our framework to describe the T3 tone sandhi domain construction in Mandarin Chinese. We hope that the variable and particular conditions of the allomorphy-related frameworks under Distributed Morphology can potentially reveal the internal logic of T3 tone sandhi in Mandarin Chinese and can also provide sufficient information about the syntactic word formation.

2.5. Summary of section 2

In section 2.1, we discussed the basic background in Distributed Morphology and the original syntactic incorporated compounding structure in Harley (2006). In 2.2, we discussed some phonological behaviors under the syntactic word formation theory. In Particular, we showed how English stress shift behavior in Marvin (2013) inspired us to adapt Shanghai Chinese stress assignment in the syntactic word structure. In 2.3, Phase theory in syntactic word formation in Embick (2010) was discussed and the concrete cyclic spell-out domains are given. Then we talked about how the revised Phase Impenetrability for Phonology in Embick (2013) determines the different sensitivities between cyclic and non-cyclic phonological rules to such phase effects. In 2.4.1, brief information in the literature about cyclic stress assignment rule in Shanghai Chinese was given and one of our core questions— the asymmetrical stress clash resolution in Shanghai Chinese was discussed with concrete examples. In 2.4.2, brief information in the literature about non-cyclic T3 tone sandhi rule in Mandarin Chinese is given. The two-pass lexical and postlexical application process in Chen (2009) for T3 tone sandhi in Mandarin Chinese was discussed with concrete examples,

[4] “M-words” means “Morphological words”

including lexical and postlexical MRU (minimal rhythm unit) domain construction. In 2.4.3, More approaches about Tone sandhi are discussed in Precompiled Phonology and Distributed Morphology. In particular, we talked about the Precompiled Rule in Hayes (1990) and Concatenation rule for domain construction in Tone sandhi in Pak (2008) and Chen (2018) under Distributed Morphology. The Concatenation rule in Pak (2008) and Chen (2018) argues that only allomorphy selection is involved after linearization of Morphological words within specific tone sandhi domain determined by morphosyntactic information (different concatenation processes of Morphological words), which is our basic framework model to develop the part of T3 tone sandhi in Mandarin Chinese.

3. Multiple-root incorporated structure of four-morpheme Chinese compounds

Three typical four-morpheme compounding structure types are taken from Chen (2009): $[[[AB]_{N1}C_{N2}]D_{N3}]/[[A_{Adj1}[BC]_{N1}]D_{N2}]/[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$, representing two main scopes of compounding structures in Chinese: N-N and A-N compounding structures. Each type is illustrated with two kinds of incorporated patterns—**a**. Consecutive pattern: multiple incorporation movements between multiple roots apply in one consecutive step. **b**. Separate pattern: multiple incorporation movements between multiple roots apply in several separate/independent steps.

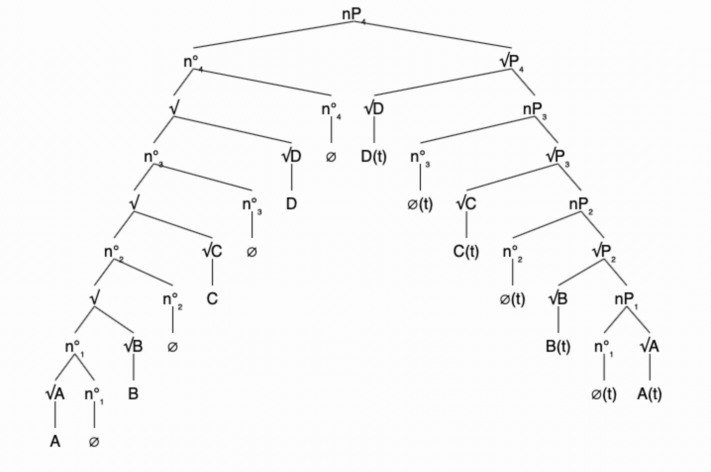
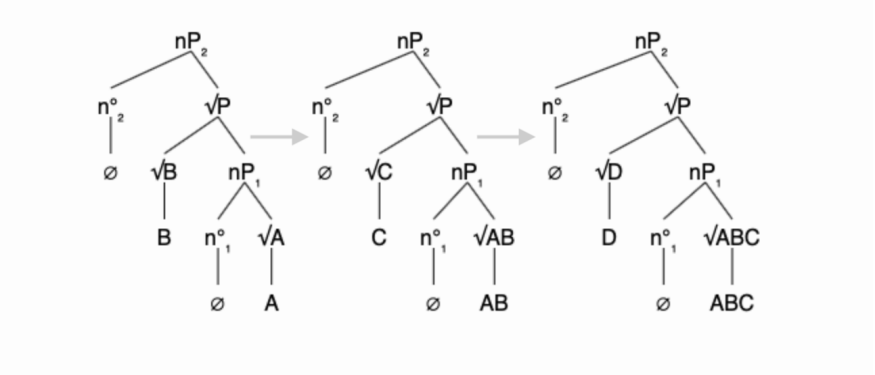
Type I $[[AB]_{N1}C_{N2}]D_{N3}$ is divided into: two-morpheme noun N1 and monomorphemic N2 and N3. Type II $[A_{Adj1}[BC]_{N1}]D_{N2}$ is divided into: two-morpheme noun N1, monomorphemic noun N2 and monomorphemic adjective Adj1. Type III $[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$ is divided into: two-morpheme noun N1, monomorphemic adjective Adj1 and monomorphemic adjective Adj2.

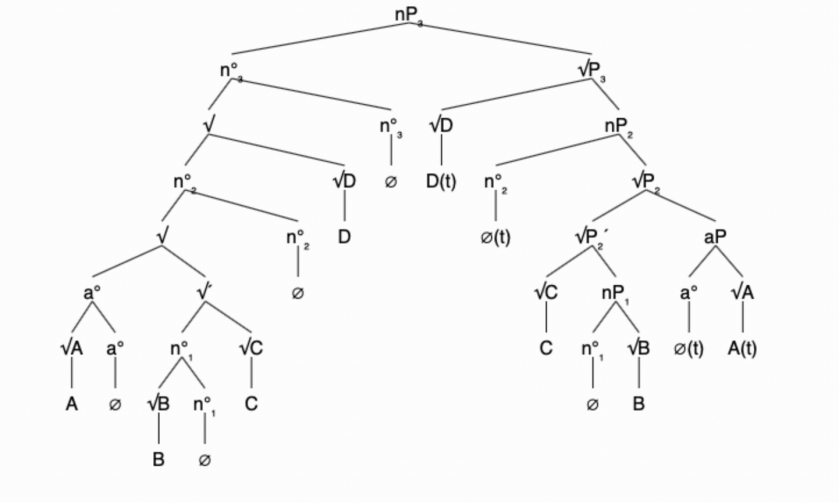
In Table 1, we are going to propose all the theoretical possibilities of incorporation movements in Consecutive pattern and Separate pattern, with the marked letters from A to F2. The concrete syntactic structures and brief explanations about different ways of incorporation movements for each subtype can be found in Table 2 with the letters corresponding from A to F2 in Table 1. All the possible multiple-root incorporated structures are adapted from the two-root incorporated compounding structure in Harley (2006) in (2). The basic logic to propose the adapted forms is to multiply the one-step incorporation

movement in Harley (2006) in (2). We assume there are two ways to multiply one single incorporation movement between only two roots: **a.** Consecutive pattern means that the multiple incorporation movements between multiple roots apply consecutively in one single syntactic tree. **b.** Separate pattern means that the multiple incorporation movements between multiple roots apply separately in different syntactic trees. We list all the possible structures of incorporation movements in an exhaustive way for each subtype in Table 1.

Table 1	Consecutive pattern	Separate pattern
Type I : $[[[AB]_{N1}C_{N2}]D_{N3}]$	A	B
Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$	C	D1 D2
Type III: $[A_{Adj1}\,[B_{Adj2}[CD]_{N1}]]$	E	F1 F2

Table 2	
Subtype	syntactic tree of possible structures
A.	Consecutive pattern for Type I $[[[AB]_{N1}C_{N2}]D_{N3}]$

	 Consecutive incorporation movements take place: √A —> √B —> √C —> √D = √ABCD
B.	Separate pattern for Type I $[[[AB]_{N1}C_{N2}]D_{N3}]$  Separate incorporation movements take place: 1. √A —> √B = √AB 2. √AB —> √C = √ABC 3. √ABC —> √D = √ABCD
C.	Consecutive pattern for Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$



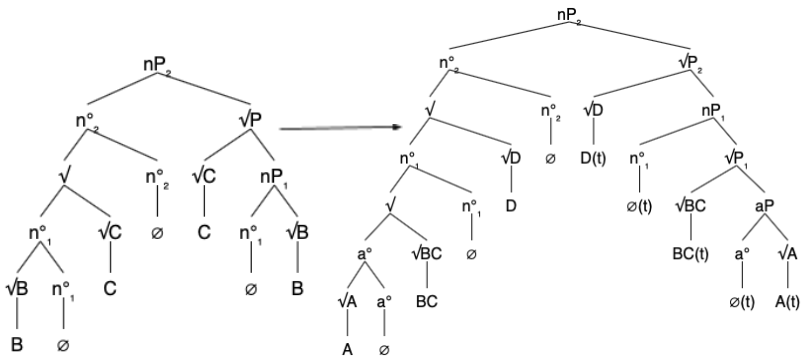
Consecutive incorporation movements take place:

√A —> (√B —>√C)—> √D = √ABCD

√A —> (√B —>√C) means that √A incorporates into the structure of combination of √B incorporating to √C.

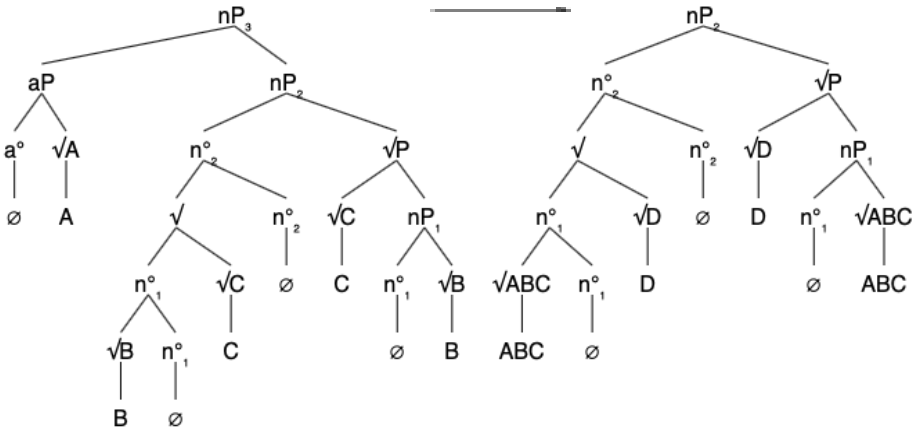
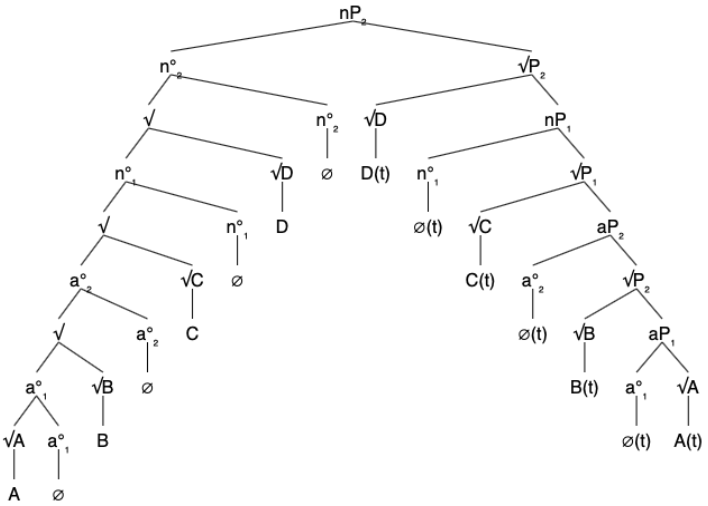
D1.

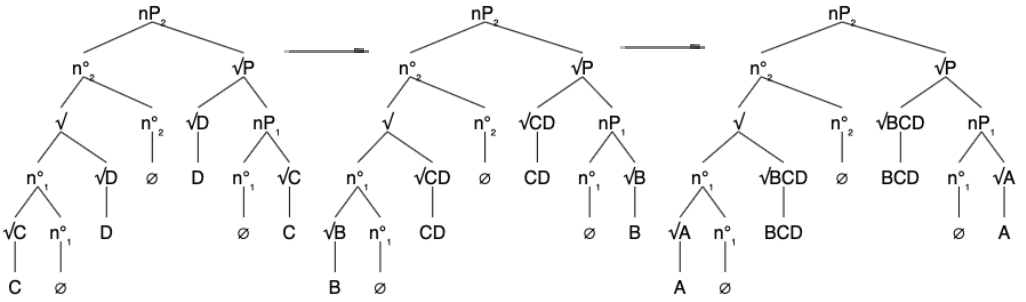
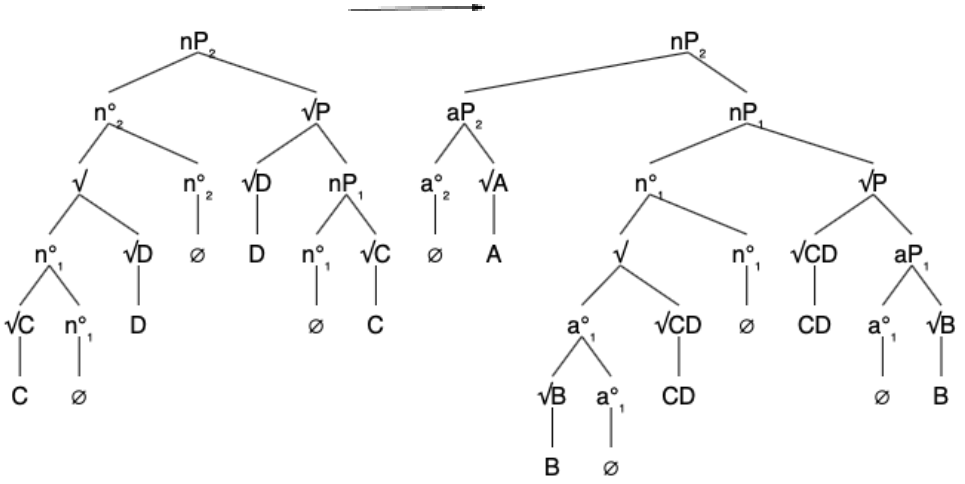
Separate pattern 1) for Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$



Separate incorporation movements take place:

1. √B —> √C = √BC 2. √A —> √BC —> √D = √ABCD

D2.	Separate pattern 2) for Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$  Separate incorporation movements take place: 1. √A ==>> (√B —>√C) = √ABC 2. √ABC —> √D = √ABCD √A ==>> (√B —>√C) means that √A is the specifier of nP3. There is no incorporation movement between √A and the structure of combination of √B incorporating to √C.
E.	Consecutive pattern for Type III: $[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$  Consecutive incorporation movements take place:

	√A —> √B —> √C —> √D = √ABCD
F1.	Separate pattern 1) for Type III: [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]]  Separate incorporation movements take place: 1. √C —> √D = √CD 2. √B —> √CD = √BCD 2. √A —> √BCD = √ABCD
F2.	Separate pattern 2) for Type III: [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]]  1. √C —> √D = √CD 2. √A ==>> (√B —>√CD) = √ABCD √A ==>> (√B —>√CD) means that √A is the specifier of nP2. There is no incorporation movement between √A and the structure of combination of √B incorporating to √CD.

To sum up, in section 3, we propose a multiple-root incorporated structure for four-morpheme Chinese compounding structure adapted from Harley (2006), and list all the possibilities of structures in Table 1 for four-morpheme Chinese compounds in two main

patterns-consecutive and separate/independent. In Table 2, we illustrate the syntactic structure with a syntax tree diagram for the structures in Table 1.

4. Stress assignment and stress resolution (stress clash avoidance) in Shanghai Chinese in multiple-root incorporated structure

4.1. Phase Impenetrability for Phonology and cyclic spell-out domains integrated with incorporated compound structure

We illustrate the Phase Impenetrability for Phonology and cyclic spell-out domains with incorporated compound structure, taking advantage of the structures proposed for $[[AB]_{N1}C_{N2}]D_{N3}$. We strictly follow the spell-out domain determined in Embick (2010), which was discussed in detail in section 2.3. For convenience, the (8) cyclic spell-out domains in the syntactic word structure determined in Embick (2010) are repeated below:

(8) The cyclic spell-out domains in the syntactic word structure, taken from Embick (2010 : 39)

a. $[[[[[\sqrt{\text{ROOT}}\ x]\ W]\ Z]\ y]$

Cyclic y triggers spell out of cyclic domains in its complement. The head x undergoes VI, as do the edge+ heads W and Z. The Root is processed phonologically.

b. $[[[[[\sqrt{\text{ROOT}}\ x^*]\ W^*]\ Z^*]\ y]\ \ldots\ z]$

Merge of higher cyclic z triggers spell out of cyclic domains in its complement. The head y defines a cyclic domain, and is subjected to VI (along with any edge+ heads it might have). The heads marked with $^*-x, W, Z-$ are present when y undergoes VI, but have undergone VI in the earlier cycle.

Taking $[[AB]_{N1}C_{N2}]D_{N3}$ as an example, the spell out domains are calculated:

$[[[[[[[\sqrt{A}\ n_1]\sqrt{B}]\ n_2\ ]\ \sqrt{C}]\ n_3]\ \sqrt{D}]\ n_4]$

Spell-out domain 1: $\sqrt{A}$ and n_1 and $\sqrt{B}$

Spell-out domain 2: $*n_1$ and $*\sqrt{B}$ and n_2 and $\sqrt{C}$ [5]

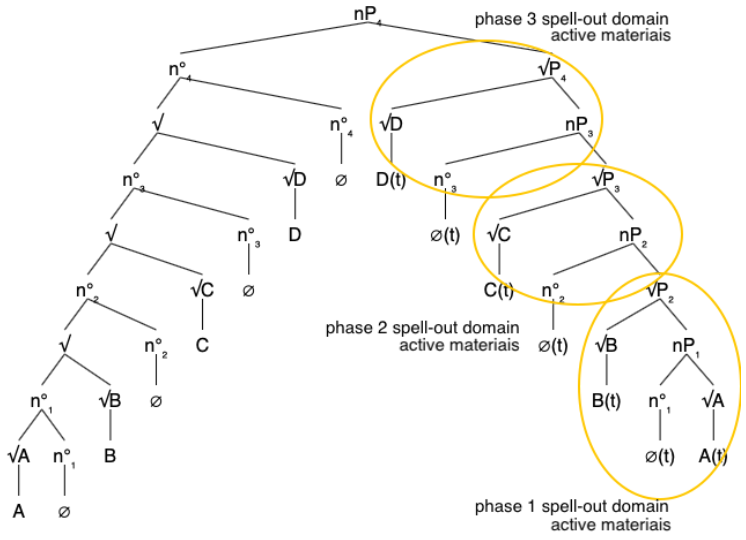
[5] Following Embick 2010, the heads marked with $*-$ n and $\sqrt{A}$ are materials with non-changeable phonological insertion from the earlier cycle. Their phonological information is present in the spell out of later cycles, but is not changeable.

Spell-out domain 3: *n_2 and *√C and n_3 and √D

or alternatively : *n_1 and *√B and *n_2 and *√C and n_3 and √D (depending on choice of proposal which is not relevant here)

In the syntax tree below, for convenience, we only illustrate the "active" materials from every spell-out domain with the "inactive" (unchangeable) materials from former spell-out domains being ruled out.

(11) "Active" materials in each spell-out domain in [[[AB[$_{N1}$C$_{N2}$]D$_{N3}$]



As we can see here in the syntax tree in (12), when we extend incorporated compounding structure to multiple-root incorporated compounding structure, e.g., the four-morpheme compounding structure [[[AB]$_{N1}$C$_{N2}$]D$_{N3}$], the spell-out domains determine that the first "A" and second root "B" belong to the same cyclic spell-out domain, which remains to be the lowest position in the syntax tree. The roots "C" and "D" belong to the separate higher cyclic spell-out domains. "A" and "B" are "inactive" when "C" is spelled out, and "A", "B" and "C" are "inactive" when "D" is spelled out. This can potentially provide new perspective to look into the uncommon stress clash avoidance between "C" and "D", as they only become "active" for once in the separate cyclic domains. As we can see in section 2.2, the English word stress shift phenomenon shows that the metrical grids are assigned synchronously under effects of cyclic spell-out phases, with former spell-out metrical grids/stresses being present but not changeable in the latter cyclic phase. In the current structure, the metrical grids assigned synchronously in the "A and B" and "C" spell-out phase, are present but not

changeable in the “D” spell-out phase. The stress clash between “C” and “D” is potentially avoided, as the metrical grids/stresses assigned to “C” are not changeable when the metrical grids/stresses assigned to “D” in the late “D” spell-out domain. In this case, the potential stress clash is structurally left over as “C spell-out domain: C-***(x)*”** + “D spell out domain: D-(x)” = C-D (x)(x)” (***(x)*** means the inactive/unchangeable /frozen metrical grids in the former spell-out domains).

4.2. Adaptation of the Shanghai stress assignment into syntactic word structure following Marvin (2013)

The stress assignment rule in Shanghai Chinese will be rewritten to combine with the syntactic word formation structure. For convenience, we repeat (1) the general stress assignment rules in Shanghai Chinese below.

(1) Shanghai Chinese Stress Assignment by Duanmu (1991, 1992a, 1993a), taken from Chen (2009 : 307)

a. Morpheme level:
Line 0: trochee, left to right, ignore degenerate foot
Line 1: left-headed, unbounded stress
b. Word/compound level: assign cyclic left-headed stress
c. Phrase level: assign cyclic right-headed stress.
d. Stress Reduction: optionally delete line 1 stress
e. Clash Resolution: remove the stress column next to a higher column

As what we talked about in section 2.2, Marvin (2013) argues that the English cyclic stress assignment rules, such as the Main Stress Rule and Edge-Marking Rules, are placed in each “xP” of syntactic word structure to naturally achieve the Cyclicity. The specific Prosodic word domain rule, e.g. Vowel reduction rule, will be carried out only after the Prosodic word is constructed. Following Marvin (2013), a similar adaptation of Stress assignment in Shanghai Chinese in the syntactic word structure is proposed as: **a**. morpheme (foot) level rule— line 0 foot construction and line 1 left-headed foot-level stress, are placed in each “xP” of syntactic word structure to naturally achieve the Cyclicity. **b**. The Word/compound level rule—cyclic left-headed stress, applies only after the Prosodic word is constructed. **c**. Phrasal level rule—cyclic right-headed stress, applies only after a Phrase is constructed. **d**. Optional Stress reduction and Clash resolution apply in the relevant situation.

Following Marvin (2013), I propose that Stress assignment rule in Shanghai Chinese can be rewritten as:

a. At every "xP", Morpheme level rule applies :
Line 0: trochee, left to right, ignore degenerate foot
Line 1: left-headed, unbounded stress
b. Word/compound level rule applies when "Prosodic word" is constructed:
assign cyclic left-headed stress
c. Phrasal level rule applies when phrase is constructed:
assign cyclic right-headed stress
d. Optional Stress reduction and Clash resolution apply in the relevant situation.

4.3. Multiple-root incorporated structure with stress assignment/stress resolution (stress clash avoidance) in Shanghai Chinese

In this section, we will use relevant data facts of stress assignment/stress resolution (stress clash avoidance) from Chen (2009) to investigate whether the theoretical possibilities proposed in section 3 may be active structures in Shanghai Chinese. We assume that not all of the theoretical structures will be supported by stress facts in Shanghai Chinese. With the cyclic domain determined in section 4.1, the adapted Stress assignment rule in Shanghai Chinese from section 4.2. will apply to each possible syntactic incorporated structure from section 3. In this way, we will later compare the output results of the stress assignment rule in Shanghai Chinese with relevant data of stress assignment from Chen (2009). If the two results are compatible, we will mark the respective syntactic incorporated structure as a successful structure. From section 4.3.1. to 4.3.3 below, we show the concrete results of the stress assignment rule in Shanghai Chinese for each type of structure from section 3.

Special focus will be given to the asymmetric stress clash case between $[[AB]_{N1}C_{N2}]D_{N3}$ —TYPE I and $[A_{Adj1}[BC]_{N1}]D_{N2}$ —TYPE II, which was discussed in detail in section 2.4.1. For convenience, we repeat the (9) asymmetric stress clash from Chen (2009) below.

(9)Asymmetrical stress clash resolution between $[[[AB]_{N1}C_{N2}]D_{N3}]$ and $[[A_{Adj1}[BC]_{N1}]D_{N2}]$, taken from Chen (2009 : 334)

Type **I**

Compound level: x
Word level: (x)(x)
Line 1: (x .)(x)(x)
Line 0: dou-fu-gan-si
"Shredded Dried Tofu"

Type **II**

Compound level: x
Word level: (x)(x)
Line 1: (x)(x .)(x)
Line 0: hu-luo-bo-si
"Shredded Carrot"

Compound level: x
Word level: (x)(x)
Line 1: (x . .)(x)
Line 0: hu-luo-bo-si
"Shredded Carrot

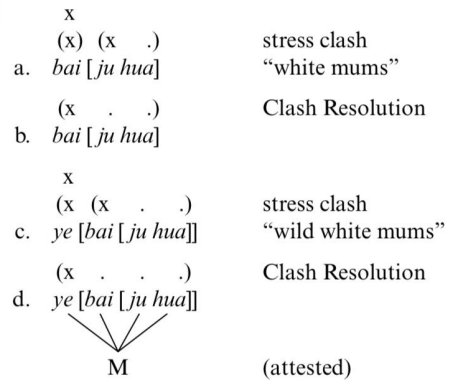
Type III [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]] will also be tested. The example (12) below from Chen (2009) shows traditional cyclicity influence in cyclic stress clash (clash resolution) in [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]]. In [ye_{Adj1} [bai_{Adj2}[ju-hua]$_{N1}$]] ("wild white mums"):

a."bai_{Adj2} (x)" and "ju-hua $_{N1}$(x .)" are constructed in the classical lexicon, where Clash Resolution Rule applies between "bai_{Adj2} (x)" and "ju-hua $_{N1}$ (x .)", forming "[bai_{Adj2}[ju-hua]$_{N1}$] (x . .)".

b. "ye_{Adj1} (x)" and "bai_{Adj2} [ju-hua]$_{N1}$ (x . .)" are constructed in the classical lexicon, where Clash Resolution Rule applies between "ye_{Adj1} (x)" and "bai_{Adj2} [ju-hua]$_{N1}$ (x . .)", forming "[ye_{Adj1} [bai_{Adj2}[ju-hua]$_{N1}$]] (x . . .)".

In Chen (2009), [ye_{Adj1} [bai_{Adj2}[ju-hua]$_{N1}$]] ("wild white mums") is used. With certain accidental reasons, we will instead use [ye_{Adj1} [bai_{Adj2}[tian-e] $_{N1}$]] ("wild white swan") in the following tests in Type III, which will not influence the test results negatively.

(12) Stress clash (clash resolution) in [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]] structure, taken from Chen (2009 : 330)

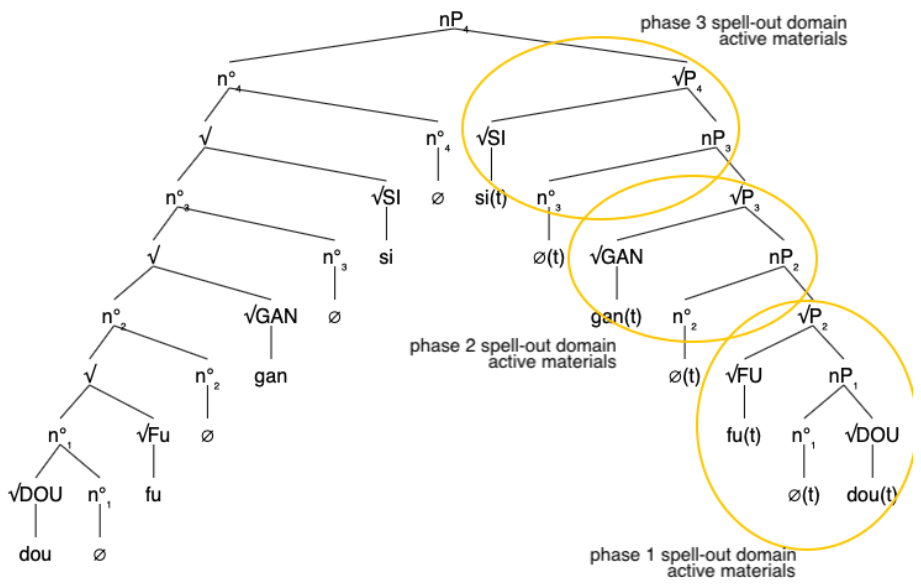


From section 4.3.1. to 4.3.3 below, we provide the concrete results of specific theoretical possibilities in Table 1 and Table 2 from Section 3. We instantiate Table 1 and Table 2 according to three different types. For example, for Type 1, we firstly instantiate the first row of Table 1 $[[[AB]_{N1}C_{N2}]D_{N3}]$ with concrete compounding structure data $[[[dou\text{-}fu]_{N1}gan_{N2}]si_{N3}]$ ($[[[tofu]_{N1}dried\text{-}food_{N2}]slice_{N3}]$). Furthermore we instantiate the syntactic trees with this data related to Type 1 in Table 2, that is A structure out of consecutive pattern and B structure out of separate pattern. For the instantiation of each relevant structure in the syntactic tree, we later illustrate the spell-out processes of cyclic phases in the steps of incorporation movements below, which yield the right or wrong results of stress assignment in Chinese Shanghai compared to the data facts from Chen (2009). Due to the separate mini-tables taken partly from Table 1 and Table 2 in section 3, we are not going to reorder these mini-tables for conciseness.

4.3.1. Type I :$[[[AB]_{N1}C_{N2}]D_{N3}]$

	consecutive pattern	separate pattern
Type I : $[[[AB]_{N1}C_{N2}]D_{N3}]$ $[[[dou\text{-}fu]_{N1}gan_{N2}]si_{N3}]$ $[[[tofu]_{N1}dried\text{-}food_{N2}]slice_{N3}]$ "Shredded Dried Tofu"	A	B

A.



A is a legitimate output of stress assignment. We illustrate the spell-out processes of cyclic phases in the steps of incorporation movements below:

1). At phase 1,

Line 1: (x .)

Line 0: dou-fu

"Tofu"

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so "dou-fu" is spelled out.

2). At phase 2,

Word level: (x)

Line 1: (x .)(x)

Line 0: dou-fu-gan

"Dried Tofu"

The complement of nP_3—the $\sqrt{P_3}$ is spelled out, so "dou-fu-gan" is spelled out.

3). At phase 3,

Compound level: x

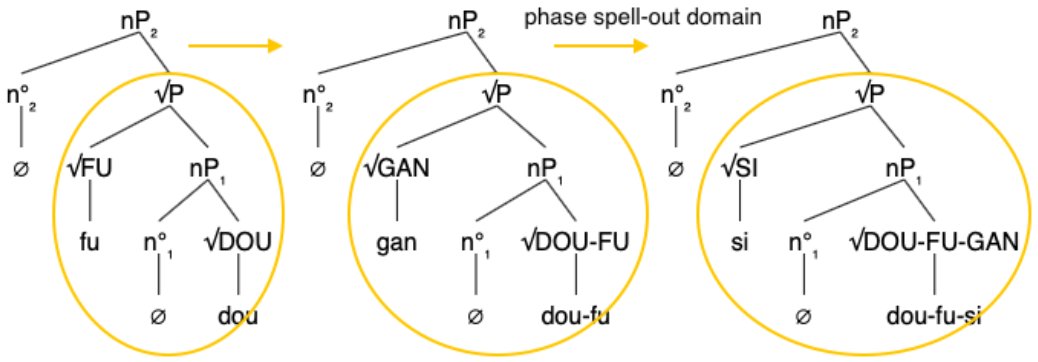
Word level: (x)(x)

Line 1: (x .)(x)(x)

Line 0: dou-fu-gan-si

"Shredded Dried Tofu"

The complement of nP_4 —the $\sqrt{P_4}$ is spelled out, so "dou-fu-gan-si" is spelled out.

B.	

By contrast, B is problematic: in the last step of incorporation above, when "√dou-fu-gan (dried tofu)" is incorporated to "√si (slice)", the spell-out domain is the √P: "√dou-fu-gan and $n_1$and √si". The stress clash (clash resolution) is imperative between "√dou-fu-gan (dried tofu) (x .)(x)" and "√si (x)", yielding a wrong output of stress assignment—"dou-fu-gan-si (shredded dried tofu) (x . .)(x)". We illustrate the spell-out processes of cyclic phases in the second step of incorporation movements below:

1). In step 1 incorporation:

Line 1: (x .)

Line 0: dou-fu

"Tofu"

The complement of nP_2 —the $\sqrt{P_2}$ is spelled out, so "dou-fu" is spelled out.

2). In step 2 incorporation:

Word level: (x)

Line 1: (x .)(x)

Line 0: dou-fu-gan

"Dried Tofu"

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so “dou-fu-gan” is spelled out.

2). In step 3 incorporation:

Compound level: x

Word level: (x)(x)

Line 1: (x .)(x)(x)

Line 0: dou-fu-gan-si

Clash resolution:

Compound level: x

Word level: (x)(x)

Line 1: (x . .)(x)

Line 0: dou-fu-gan-si

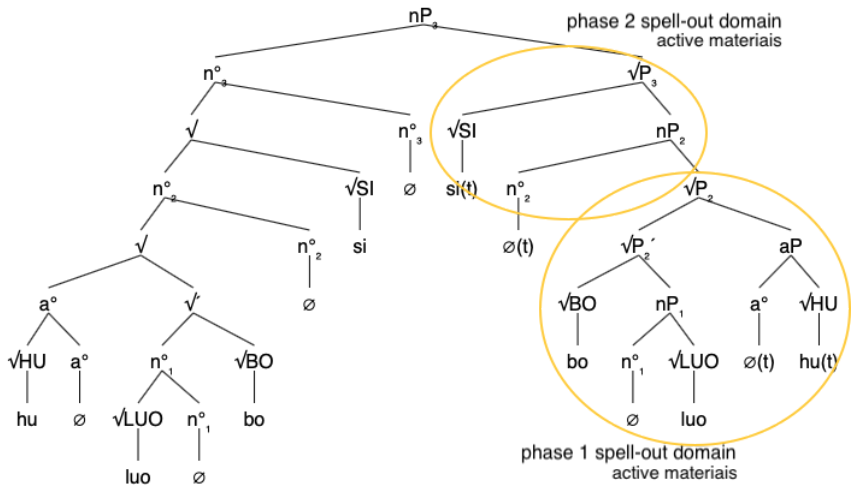
“Shredded Dried Tofu”

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so “dou-fu-gan-si” is spelled out.

4.3.2. Type II :$[[A_{Adj1}[BC]_{N1}]D_{N2}]$

	consecutive pattern	separate pattern
Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$ $[[hu_{Adj1}[luo\text{-}bo]_{N1}]si_{N2}]$ $[[foreign_{Adj1}[raddish]_{N1}]slice_{N2}]$ “Shredded Carrot”	C	D1 D2

C.



C is also problematic: With the foot construction determined in foot-level stress (Line 0: trochee, left to right, ignore degenerate foot), in spell-out domain of phase 1, the complement of nP_2— the $\sqrt{P_2}$—"√hu-luo-bo (carrot)". The line 1 stress (metrical grids) is "√hu-luo-bo (carrot) (x .)(x)", which is a wrong output. We illustrate the spell-out processes of cyclic phases in the steps of incorporation movements below:

1). At phase 1,

Word level: (x)

Line 1: (x .)(x)

Line 0: hu-luo-bo

"Carrot"

The complement of nP_2 the $\sqrt{P_2}$ is spelled out, "hu-luo-bo" is spelled out.

2). At phase 2,

Compound level: x

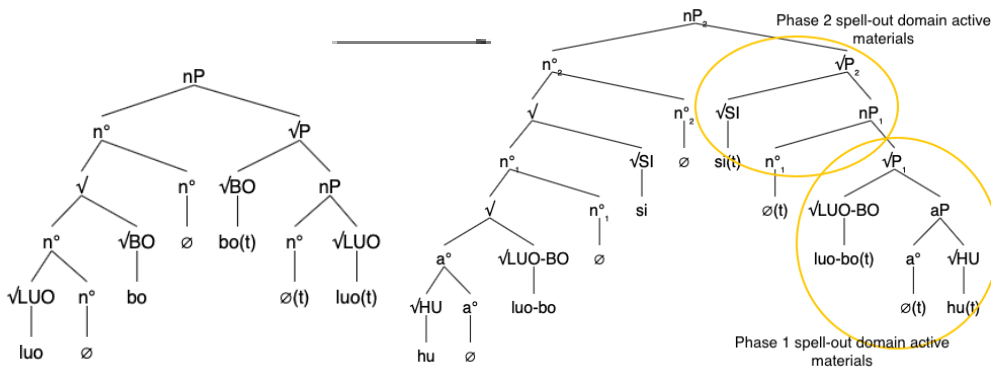
Word level: (x)(x)

Line 1: (x .)(x)(x)

Line 0: hu-luo-bo-si

"Shredded Carrot"

The complement of nP_3 —the $\sqrt{P_3}$ is spelled out, so "hu-luo-bo-si" is spelled out.

D1.	

D1 is a legitimate option of stress assignment result of [[A_{Adj1}[BC]$_{N1}$]D_{N2}]- [[hu_{Adj1}[luo-bo]$_{N1}$]si_{N2}]. In the first step of incorporation, "√luo" is incorporated into "√bo" to form a new "√luo-bo (raddish)" root. In the first step of incorporation, the stress assignment rule is applied in the basic regular form: from left to right, "√luo-bo (raddish)" forms the trochee foot with metrical grids "√luo-bo (raddish) (x .)". We illustrate the spell-out processes of cyclic phase 2 in the steps of incorporation movements below:

1). At phase 1,

Word level: (x)

Line 1: (x)(x .)

Line 0: hu-luo-bo

"Carrot"

Clash resolution:

Word level: (x)

Line 1: (x . .)

Line 0: hu-luo-bo

"Carrot"

The complement of nP_1—the $√P_1$ is spelled out, so "hu-luo-bo" is spelled out.

2). At phase 2,

Compound level: x

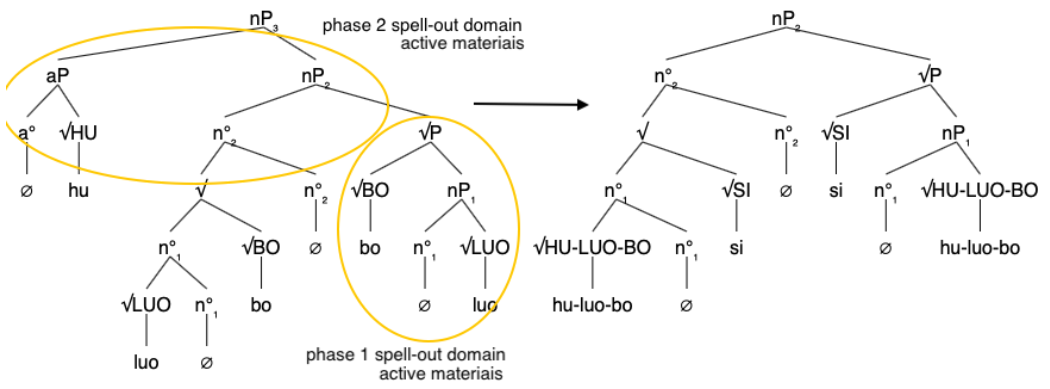
Word level: (x)(x)

Line 1: (x . .)(x)

Line 0: hu-luo-bo-si

"Shredded Carrot"

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so "hu-luo-bo-si" is spelled out.

D2.	

Unlike D1, D2 is problematic: at phase 2, the Clash Resolution Rule between "√hu (foreign) (x)" and "√luo-bo (raddish) (x .)" is blocked, due to the "√luo-bo (raddish) (x .)" is unchangeable spell-out material from phase 1, yielding a wrong result.

In step 1 incorporation:

1). At phase 1,

Line 1: (x .)

Line 0: luo-bo

"Raddish"

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so "luo-bo" is spelled out.

2). At phase 2,

Word level: (x)

Line 1: (x)(x .)

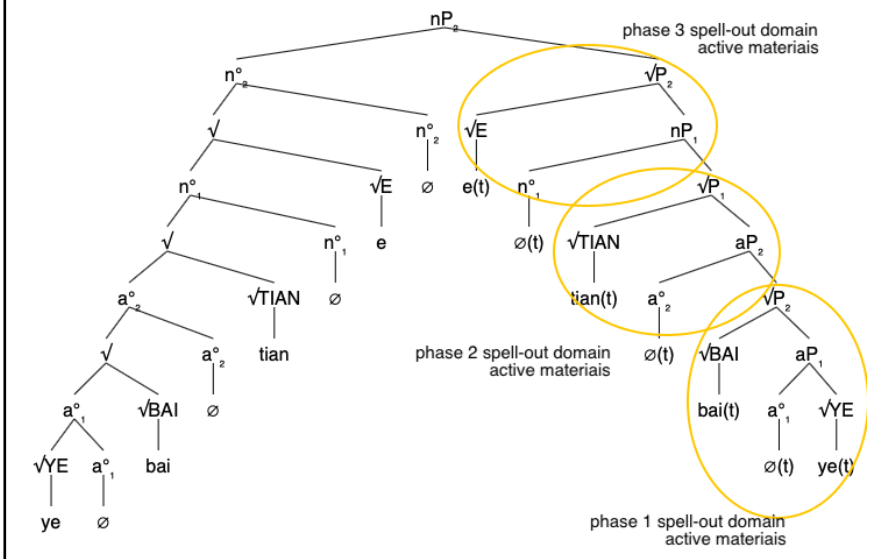
Line 0: hu-luo-bo

"Carrot"

Clash resolution is not applied due to “luo-bo (raddish) (x .)” being the phase 1 unchangeable spell-out material. The complement of a higher phrase nP_3—the aP and nP_3 are spelled out, “hu-luo-bo (carrot) ” is spelled out.

4.3.3. Type III :[A_{Adj1} [B_{Adj2}[CD] $_{N1}$]]

	consecutive pattern	separate pattern
Type III: [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]] [ye_{Adj1} [bai_{Adj2}[tian-e] $_{N1}$]] [$wild_{Adj1}$ [$white_{Adj2}$[swan] $_{N1}$]] “ wild white swan”	E	F1 F2

E.	

E is problematic: the consecutive incorporation movements take place from √ye —> √bai —> √tian —> √e, yielding a wrong result of metrical grids (x .)(x)(x). We illustrate the spell-out processes of cyclic phases in the steps of incorporation movements below:

1). At phase 1,

Line 1: (x .)

Line 0: ye-bai

“wild-white”

the complement of aP_2—the $\sqrt{P_2}$ is spelled out, so “ ye-bai” is spelled out.

2). At phase 2,

Word level: (x)

Line 1: (x .)(x)

Line 0: ye-bai-tian

“wild-white-tian(sky)”

the complement of nP_1—the $\sqrt{P_1}$ is spelled out, so “ ye-bai-tian” is spelled out

3). At phase 3,

Compound level: x

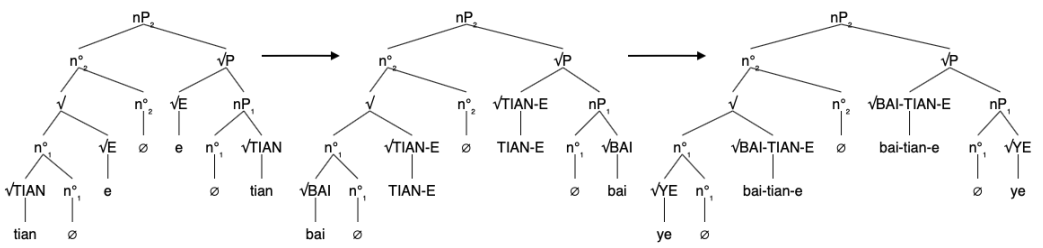
Word level: (x)(x)

Line 1: (x .)(x)(x)

Line 0: ye-bai-tian-e

“wild white swan”

the complement of nP_2—the $\sqrt{P_2}$ is spelled out, so “ye-bai-tian-e” is spelled out.

F1	

F1 is a legitimate option of stress assignment of $[A_{Adj1}\,[B_{Adj2}[CD]_{N1}]]$-$[ye_{Adj1}\,[bai_{Adj2}[tian\text{-}e]_{N1}]]$. We illustrate the spell-out processes of cyclic phases in the steps of incorporation movements below:

1). In step 1 incorporation,

Line 1: (x .)

Line 0: tian-e

“swan”

the complement of nP_2—the $\sqrt{P_2}$ is spelled out, so “ tian-e” is spelled out.

2). In step 2 incorporation,

Word level: (x)

Line 1: (x)(x .)

Line 0: bai-tian-e

Clash resolution:

Word level: (x)

Line 1: (x . .)

Line 0: bai-tian-e

“white swan”

the complement of nP—the $\sqrt{P}$ is spelled out, so “ bai-tian-e” is spelled out.

3). In step 3 incorporation,

Compound level: x

Word level: (x)(x)

Line 1: (x)(x . .)

Line 0: ye-bai-tian-e

Clash resolution:

Compound level: x

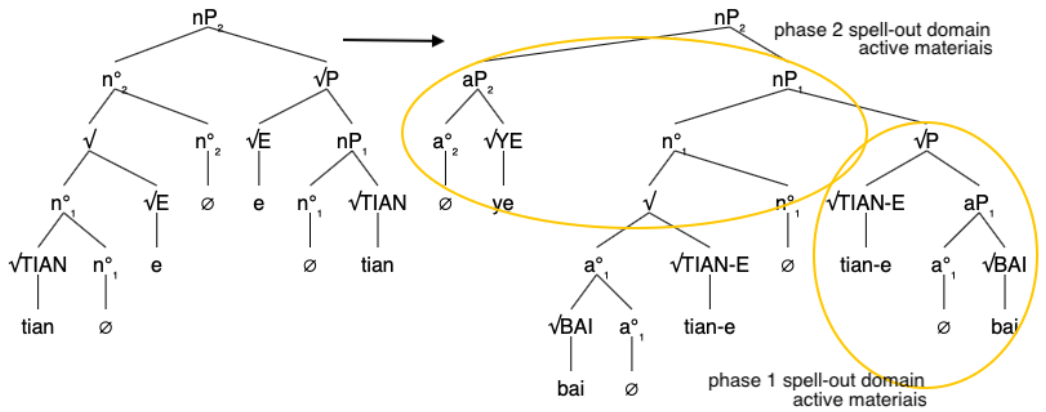
Word level: (x)(x)

Line 1: (x . . .)

Line 0: ye-bai-tian-e

“wild white swan”

The complement of nP_2—the $\sqrt{P_2}$ is spelled out, so “ye-bai-tian-e” is spelled out.

F2	

Finally, F2 is problematic: in the second step of incorporation, at phase 2, line 1 foot level stress of "√bai-tian-e (white swan) (x . .)" in phase 1 is unchangeable in the subsequent phases, which blocks the imperative stress clash between and "√ye (wild) (x)" and "√bai-tian-e (white swan) (x . .)" in phase 2, yielding a wrong output.

We illustrate the spell-out processes of cyclic phases below:

1). In step 1 incorporation,

Line 1: (x .)

Line 0: tian-e

"swan"

the complement of nP_2 —the $\sqrt{P_2}$ is spelled out, so "tian-e" is spelled out.

2). In step 2 incorporation,

a.At phase 1:

Word level: (x)

Line 1: (x)(x .)

Line 0: bai-tian-e

Clash resolution:

Word level: (x)

Line 1: (x . .)

Line 0: bai-tian-e

“white swan”

the complement of nP— the √P is spelled out, so “ bai-tian-e” is spelled out.

b.At phase 2:

Compound level: x

Word level: (x)(x)

Line 1: (x)(x . .)

Line 0: ye-bai-tian-e

Clash resolution is not applied due to “√bai-tian-e (white swan) (x . .)” being the phase 1 unchangeable spell-out material. The complement of a higher phrase—nP_2 —the aP_2 and nP_1 are spelled out and “ye-bai-tian-e (wild white swan) (x)(x . .)” is spelled out.

4.5. Brief conclusion about multiple-root compounding structure with Shanghai Chinese stress assignment data

With the analysis above, we have ruled out the unsuccessful options proposed in the theoretically possible multiple-root incorporated structure for the four-morpheme Chinese compounding structures in section 3, and the successfully tested results are marked in red in the table below.

The multiple-root incorporated structure of four-morpheme Chinese compounds are shown to be divided into two patterns—Consecutive pattern in Noun-Noun compounds ($[[[AB]_{N1}C_{N2}]D_{N3}]$) and Separate pattern in Adjective-Noun compounds ($[[A_{Adj1}[BC]_{N1}]D_{N2}]$ and $[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$).

Table 1	consecutive pattern	separate pattern
Type I : $[[[AB]_{N1}C_{N2}]D_{N3}]$	A	B
Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$	C	D1 D2
Type III: $[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$	E	F1 F2

4.6. Summary of section 4

In section 4, we firstly integrate Phase Impenetrability for Phonology/cyclic spell-out domains from Embick (2010 ; 2013) with the multiple-root incorporated structure, then we rewrite and adapt the Shanghai Stress assignment rule in the syntactic word formation following Marvin (2013), and finally we check each type of multiple-root incorporated structure of four-morpheme compounding structure in Mandarin Chinese in section 3 with the data of Shanghai Chinese stress assignment rule and draw a conclusion to rule out the unsuccessful results.

5. T3 tone sandhi domain construction as Concatenation rule in Mandarin Chinese

5.1. Criterion under consideration in allomorphy-related proposals

Pak (2008) proposed the head-left and phrasal-left Concatenation processes, which determine the locality features between morphemes: **a**. Head-left Concatenation: identifies pairs of M-words X, Y where (i) X is left-adjacent to Y, and (ii) X c-commands Y. **b**. Phrase-left Concatenation: identifies pairs of M-words X, Y where (i) X is left-adjacent to Y, and (ii) X does not c-command Y. Pak (2008) also describes that "Chaozhou tone sandhi applies strictly from right to left across a string of words within a given domain, regardless of the syntactic bracketing." We will check if we can construct an algorithm related to the two kinds of concatenation processes to describe the morphosyntactic relations in the data of T3 tone sandhi domain construction in Mandarin Chinese from Chen (2009).

Generally, inspired by the literature, we are investigating the following questions about T3 Tone sandhi in Mandarin Chinese:

a. How the Morphological words are linearized adjacent to each other in the two kinds of concatenation process—head-left and phrasal-left Concatenation. **b**. Whether phonological insertion (vocabulary insertion) has access to non-local information about syntactic structure, e.g., c-command relationship between Morphological words (Pak 2008). **c**. How the word/compound formation process can be integrated into the sandhi rule application in T3

tone sandhi in Mandarin Chinese. **d**. How the spell-out phases can influence the whole process.

In Table 3, we list the abbreviations to refer to the indications that we will use to construct the algorithm for the domain construction of Tone sandhi in Mandarin Chinese. The main morphosyntactic relations under consideration are: **a.** c-command relations **b.** locality features of Morphological words (phrasal category or non-phrasal category)

Table 3	***P**: morphological word in locality of complex internal syntactic structure (e.g. N in NP in concatenation with subsequent V)* ***M**: morphological word in locality of simple internal syntactic structure (e.g. V in VP in concatenation with subsequent Object NP. The locality of V can only be limited to the head position of VP instead of entire complete VP)*(an example to illustrate the difference between P and M following this chart)* *#: T3 sandhi rule domain boundary. T3 # T3 means the latter T3 cannot trigger the former T3 to T2* ***A B c.**: Morphological word A c-commands B* ***A B n.**: Morphological word A doesn't c-command B*

an example from Chen (2018) to illustrate the difference between "P**"and "**M**": In (13) below, "Assume that 'kah (and)' or an empty operator is the head of &P, and that the first AP 'khoai-lok (happy)' is the specifier of the &P while the second AP "peng-cheng (quiet)" is the complement, the head of &P 'kah (and)' is not a phrasal category." So we can generate a syntactic tree with &P below, the head of the &P—'kah (and)' is not a phrasal category, which means it belongs to "**M**" locality instead of "**P**".*

(13)

in koe-tioh chin khoai-lok # (kah) peng-cheng # e seng-oah #
They live very happy and quiet E life
'They lead a very happy and quiet life.'

5.2. Results of concatenation process of data at the sentence level in Chen (2009)

Our analysis of T3 Mandarin tone sandhi is given in Table 4 below. Most of the examples of tone sandhi are from Chen (2009), and some restricted adaptations of these examples are also provided out of intuition from native speakers. About this part of data, the adapted forms from the original data from Chen (2009) are under strict control. With limited resources, testing the authenticity to some extent that is more general than corpus in Chen (2009) is currently not the goal here.

In Table 4, for each example of the data in Chen (2009), we list: **a).** the concrete syntactic structure. **b).** the tone sandhi results from the MRU domain construction process in Chen (2009) with page number (for each example: the first row of tones are basetones and the following rows of tones are sandhi forms, with number 2 marking the sandhi form T2 and number 3 marking the citation form T3). **c).** the algorithm of concatenation results based on **a)** and marked with symbols we denominated in Table 3. For example, in the first row **1a,** the syntactic structure **[$_{VP}$xie[$_{NP}$xiaoshuo]]** is shown in the first column with respective English translation **[$_{VP}$write[$_{NP}$novel]]**. The Tone sandhi results in the framework from Chen (2009) with specific page numbers are shown in the second column—**(s 3 1)**. The concatenation results based on the discussion in 5.1.—**(M P c.#)** is given in the third column. As what we have discussed in section 5.1, **(M P c.#)** means M structure c-commands P structure and there is a natural sandhi boundary at the end of the structure. We use red to mark Head-left Concatenation (M …P c.#) and blue to mark Phrase-left Concatenation (P…M n.#). Detailed explanations about each example of data are given after Table 4.

Table 4	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
1a.	[$_{VP}$ xie [$_{NP}$ xiaoshuo]] [$_{VP}$ write [$_{NP}$ novel]]	write novel xie [xiao-shuo] 3 3 1 (s 3 1)	M P c.#

		p377	
1b.	**1b1** [$_{CP}$ gou [$_{VP}$ yao wo]] [$_{CP}$ dog [$_{VP}$ bite me]] **1b2** [$_{CP}$ wo [$_{VP}$ xie [$_{NP}$ shu]] [$_{CP}$ I [$_{VP}$ write [$_{NP}$ book]] **1b3** [$_{CP}$ gou [$_{VP}$ yao [$_{NP}$ nü-ren]]] [$_{CP}$ dog [$_{VP}$ bite [$_{NP}$ woman]]]	**1b1** dog bite me gou [yao wo] (3 (3 3)) (s 3) TS (3 s 3) TS not applicable = reading (i) --- - ((3 3) 3) Alternative Foot Formation (s 3) TS (s s 3) TS = reading (j) p418 **1b2** I write book wo [xie shu] 3 3 1 (s 3 1) p377 **1b3** “the dog bit the woman” [$_{CP}$ gou [$_{VP}$ yao [$_{NP}$ nü-ren]]] (3 3 3 2) (2 3) (3 2) p411	**1b1** P # [M n. P c.] **1b2** [P M n.]? P c. **1b3** [P M n.]# P c.
2.	[$_{VP}$ xiang [$_{CP}$ pro mai gudong #]]	want buy antique	M M c P c.#

	[$_{VP}$ want [$_{CP}$ pro buy antiques #]] "(null subject) want to buy antiques " [$_{VP}$ xiang [$_{CP}$ pro xie xiaoshuo #]] [$_{VP}$ want [$_{CP}$ pro write novel #]] "(null subject) want to write novel" (in Chinese there is no "to" after "want" to occupy the T position in TP.)	"wants to buy antiques" xiang [mai [gu-dong]] 3 3 3 3 (2 3) (2 3) want write novel xiang # [xie # xiao-shuo] (3 3 3 1) (s 3) (3 1) Phrasal MRU, TS i = ok (s s) (3 1) Cross-MRU TS optional j = ok p386	
3.	[$_{CP}$ wo [$_{VP}$ xiang # [$_{CP}$ pro mai gupiao #]]] [$_{CP}$ I [$_{VP}$ want # [$_{CP}$ pro buy stocks #]]]		P M n. # M c P c. #
4.	[$_{CP}$ wo [$_{VP}$ xiang # [$_{CP}$ pro mai shu #]]] [$_{CP}$ I $_{VP}$[want # [$_{CP}$ pro buy book#]]]	I plan buy book wo # [xiang # [mai # shu]] 3 3 3 1 (s 3) (3 1) p381	P M n. # M c P c #
5.	[$_{VP}$ mai [$_{NP}$ [$_{PP}$ gei ni] # gupiao]]]] [$_{VP}$ buy [$_{NP}$ [$_{PP}$ for you] # stocks]]] "buy you stocks"		M M c. P c. # P n.

	(in Chinese "for" here is necessary when an indirect object is inverted with a direct object.)		
6.	[$_{VP}$ mai [$_{NP}$ gupiao [$_{PP}$ gei ni #]]] [$_{VP}$ buy [$_{NP}$ stocks [$_{PP}$ for you #]]]		M M c. M c. P c. #
7.	[$_{CP}$ wo $_{VP}$[xiang # [$_{CP}$ PRO [$_{VP}$ mai [$_{NP}$ gupiao#[$_{PP}$ gei ni#]]]]]] [$_{CP}$ I $_{VP}$[want # [$_{CP}$ PRO [$_{VP}$ buy [$_{NP}$ stocks # [$_{PP}$ for you#]]]]]]		P M n. # M c. M c.? M n. P c.#
8.	[$_{baP}$ ba [$_{VP}$ [$_{NP}$ bi] # gei Xiaoming #]]] [$_{baP}$ ba [$_{VP}$ [$_{NP}$ pen] # give Xiaoming #]]] "give you the pen" (ba structure: ba +direct object +verb+indirect object =verb + direct and indirect objects)	Xiaomei BA pen give Xiaoming "Xiaomei gave the pen to Xiaoming" [xiao-mei] [[ba bi] [gei xiao-ming]] 3 3 3 3 3 3 2 (2 3) (2 3) (2 3 2) p407	M P c.# M n. P c.#
9.	[$_{baP}$ ba [$_{VP}$ [$_{NP}$ bi] # zhao [$_{PP}$ gei ni #]]] [$_{baP}$ ba [$_{VP}$ [$_{NP}$ pen] # find [$_{PP}$ for you #]]] "find the pen for you" (ba structure: ba +direct object +verb+indirect object =verb + direct and indirect objects)		M P c.# M n. M c. P c. #
10.	[$_{CP}$ zhi-laohu [$_{VP}$ pao]] # [$_{CP}$ paper tiger [$_{VP}$ runs]] #	the paper tiger runs [zhi-[lao-hu]] pao 3 3 3 3	P M n.#

		(3 s s 3) p373	
11.	[$_{VP}$ [$_{ADVP1}$[$_{ADVP2}$ hen] zao] qi] # [$_{VP}$ [$_{ADVP1}$ [$_{ADVP2}$ very] early] get up] # “ get up very early” ADVP2 here is the degree adverb phrase to describe “early”	very early rise [hen zao] qi 3 3 3 (s s 3) p376	P M n. M n.#
12.	[$_{CP}$ gou [$_{VP1}$ chao **xing** # [$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] [$_{CP}$ dog [$_{VP1}$ barks **wakes up** #[$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] “the dog barks so that Xiaomei wakes up” Verb incorporation Structure of verb + resultative verb	dog noisy-wake up “the dog woke up Xiaoming by making noise” gou $_{VP}$[$_{V}$[chao-xing] $_{NP}$[xiao-ming]] 3 3 3 3 2 (s 3) (3 2) Lexical MRU, TS (3 s 3) (3 2) Postlexical MRU, TS not applicable (optional between MRUs) p391	P M n. # P n.

The table above is arranged by syntactic structure from simple to complex scale.

1a is simply a VO structure. Recalling the definition of phrasal category and non-phrasal category—locality of complex and simple internal syntactic structure, the object NP acts as the complement of VP, so the head N belongs to the phrasal category as the head V is a non-phrasal category structure, which means M c-commands P structure without definite tone sandhi structure.

1b is SVO sentence structure. The subject NP is the specifier of higher IP or TP, whose head N belongs to phrasal category locality. Between the subject and verb, the head N in subject does not c-command the head V—P does not c-command M. Between the verb and object, the process is the same as the **1a**. The whole structure is P M n. P c. However, according to Chen (2009), the TS boundaries **1b1**, **1b2**, **1b3** are different with the same syntactic structure. Apart from the object NP in **1b3** is disyllabic, these 3 sentences share the same structure in syntax. With obvious different results—**1b1** shows TS boundary between the subject and verb; **1b2** shows no TS boundary between the subject and verb and potential TS boundary between verb and object; **1b3** shows no TS boundary between the subject and verb but definite boundary between verb and object. The results will be discussed later, which will show the two-morpheme object in **1b3** can be very useful to discover the authentic logic behind the controversial TS behaviors here.

2 is (null subject) V + Clause (finite clause), which is intended to examine the phase theory effects following Embick (2013). The finite clause itself is still similar to **1a** that is a VO structure—M c-commands P. The matrix verb V and the following verb V from the finite clause both belong to the non-phrasal category, which means M c-commands M. The whole structure is M M c. P c. and there is no obvious TS boundary between them.

3and 4 is S + V + Clause (finite clause). The V + finite clause part is similar to **2**, which is M M c P c. The subject is an independent NP as the head N from the subject belongs to the phrasal category, while the head V belongs to the non-phrasal category, which constructs P M n. # M c. P c. and TS boundary is between the matrix V and V from finite clauses.

567 is about the analysis of prepositional phrase structure inspired by **8** from Chen (2009). Due to the flexibility of prepositional phrase position in Chinese, we have **5—**verb + PP + NP, which entails a PP as a specifier of the NP. The preposition head of prepositional phrase is c-commanded by the former V and at the same time c-commands the latter N from the internal NP of PP. **6** is verb + PP + NP, which PP takes the complement position of NP. The head V c-commands head P while head P doesn't c-commands the head N. **7** is combined with the finite clause. **567** are from the intuition of native speakers, so this part will not be discussed too much. We view this part as some potential inspired evidence from Chen (2009).

89 is a special ba-sentence structure- ba +direct object +verb+indirect object =verb + direct and indirect objects. In **8** the baP takes the direct object NP as complement which takes VP

containing internal indirect object NP as complement. The head “ba” c-commands the head N from direct object NP which c-commands the head V which c-commands the head N of the indirect object NP. Apart from the head N of the indirect object NP belongs to the phrasal category, the other heads above belong to the non-phrasal category. In **9** the basic baP structure is the same but with additional PP above the indirect object NP.

11 is the AdvP structure: AdvP 1 is the Vp specifier and has internal degree AdvP 2 as its own specifier. AdvP2 is an independent structure belonging to phrasal category, the Adv1 head is a non-phrasal category and the external V head from VP is non-phrasal category, yielding a whole structure as (P M n. M n.) respectively.

12 is v + resultative verb VV structure. Following Chen (2009), the resultative verb is considered to be incorporated into the action verb. Chen (2009) treated this special incorporated structure as an exception to be lexical MRU domain, undergoing the first step lexical tone sandhi as other common lexical words. However, from native speakers' intuition, “verb + resultative verb” VV structure can hardly be a lexical word compared to other common lexical words. Even though lexical domain may not be the key criteria, the incorporated structure can potentially be the reason why this structure shows a more intimate relation with lexical domain. The specialty of resultative structure may not be related to whether they are lexical or not, alternatively may be because they share the same structure of incorporated compounding structure (v + resultative verb structure is proposed to be constructed by incorporation movement in the literature).

5.3. Brief analysis about data at the sentence level

Overall, from the table above:

1. Spell-out phases are invisible in tone sandhi domain, which means one sandhi domain can include Morphological words from different spell-out phases. In **2** from Table 4, the matrix verb and the non-finite clause are integrated into one single sandhi domain. This could conform to Embick (2010 and 2013) that “inactivity” features of materials from former cyclic spell out domains “can only be seen and altered by non-cyclic or phrasal phonological rules, but not by cyclic phonological rules”.

2. While we didn't expect, the results of sandhi domains are highly-uniformed, divided into two types : (P...M n.) and (M...P c.), which conform to the Phrasal-left concatenation rule and Head-left concatenation rule in Pak (2008), respectively.

From the perspective of morphological concatenation process, even though the phonological behaviors of tone sandhi across Chinese dialects are well defined as variant and distinctive, the temporary conclusion of T3 tone sandhi rule is really similar to the sandhi domain construction rule about Min-Taiwanese proposed by Chen (2018) : "Tone sandhi domains in TSM are determined by a Head-left Concatenation rule (Pak 2008), a Chaining operation takes place to establish linear order across all the Morphological words between a pair of sandhi domain boundaries, and lastly, the phonological context-free (tone sandhi) rule applies from left to right within that chain of Morphological words."

5.4. Brief framework of T3 Tone sandhi domain construction in Mandarin Chinese as Concatenation rule at the sentence level

Adapted from Chen (2018), we temporarily conclude the framework of T3 Tone sandhi domain construction in Mandarin Chinese as Concatenation rule at the sentence level:

1) Temporarily with the limited data available, we assume the T3 tone sandhi rule can be phase-free, meaning that the spell-out phases have little effect on the sandhi domain distribution in T3 tone sandhi in Mandarin Chinese.
2) In each sentence, **M** structure from left to right searches for the possibility of Head-left Concatenation (illustrated in M...P c. in Table 4): identifies pairs of M-words X, Y where (i) X is left-adjacent to Y, and (ii) X c-commands Y, and **P** structure from left to right searches for the possibility of Phrase-left Concatenation (illustrated by P...M n. in Table 4), on the other hand , identifies pairs of M-words X, Y where (i) X is left-adjacent to Y, and (ii) X does not c-command Y. (**M** and **P** are syntactic structure locality determined by containing Morphological words)
3) For Concatenation statements continues as A⌢B⌢C…, a sandhi domain boundary is inserted when no further Concatenation statement exists, in other words a sandhi domain boundary is inserted after the ultimate Morphological word which can comply with current Concatenation algorithm.
4) Once one sandhi domain is determined, the subsequent new concatenation process is continued from the next Morphological word, repeating the process in 2), until the

end of the sentence. There is one exception to this, in example 1b above: **1b1** P # [M n. P c.]; **1b2** [P M n.]? P c.; **1b3** [P M n.]# P c. The sequence of (P M n. P c.) is controversial in determining the sandhi domain boundary. From **1b1** to **1b3**, there is no clear evidence to explain the logic that competition between M...P c. algorithm (Head-left Concatenation) and P...M n. algorithm (Phrase-left Concatenation). However, the difference between **1b1** and **1b3** in the root number of N from internal NP of VP (monomorphemic and disyllabic), may provide important information about this, which also is identified when word formation is taken into consideration in this Concatenation process. See the discussion about the word formation part in section 6.2 below.

5) Tone sandhi rule as Chaining operation takes place to establish linear order across all the Morphological words, from left to right within that chain of Morphological words in every sandhi domain.

6) Following Chen (2018), we determine the tone sandhi rule in Mandarin Chinese changing from citation form T3 to sandhi form T2: T3—> T2 / ___T3.

5.5. Summary of section 5

In section 5, we propose the new framework of T3 tone sandhi domain construction as the Concatenation rule in Mandarin Chinese, and formulate the tone sandhi domains determined by Morphological word concatenation process (morphosyntactic information—morphosyntactic locality characteristics and c-command relations). In part 5.1, we discuss the criterion listed in Table 2 in allomorphy-related proposals in the literature, which we use to formulate Morphological words concatenation process to construct the tone sandhi domains where the framework of Concatenation rule applies. In 5.2, we give the results of the concatenation process with the data from Chen (2009) in T3 tone sandhi in Mandarin Chinese at the sentence level in Table 3, with detailed information about how the different syntax structures influence the corresponding criterion and lead to variable concatenation results. In 5.3, a quick analysis about the data at the sentence level is given, confirming that: a. Cyclic spell-out phases are invisible in tone sandhi domain in Mandarin Chinese b. Concatenation results of tone sandhi domains from the data in Mandarin Chinese are highly-uniformed, divided into two types: the Phrasal-left concatenation rule and Head-left concatenation rule in Pak (2008). In 5.4, the brief framework of T3 Tone sandhi

domain construction in Mandarin Chinese as Concatenation rule at the sentence level is given.

6. T3 tone sandhi domain as Concatenation rule with multiple-root incorporated structure

In this section, we are going to add the successful multiple-root incorporated structures from section 3 and 4 into the concatenation results marked with algorithms. The examples with three-morpheme compounding structures and four-morpheme compounding structures will be discussed respectively.

6.1. Three-morpheme compounding structures

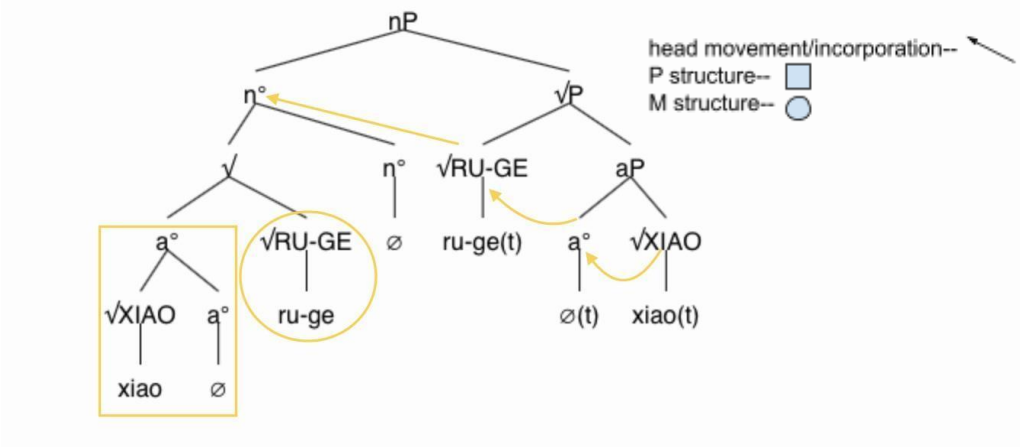
We start from three-morpheme compounding structures to see how incorporated compounding structures can be integrated into the Concatenation process regulated above. To check the interface below and above the classical word domain, the $A_{Adj}[BC]_{N.}$ and $A_{N1}[BC]_{N2}$ in VP structure will be discussed in Table 5. The incorporation pattern of $A_{Adj}[BC]_{N.}$ and $A_{N1}[BC]_{N2}$ comes from the successful D1 structure for $[[A_{Adj1}[BC]_{N1}]D_{N2}]$.

Table 5	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
13.	[$_{VP}$[$_{NP}$ zhi [laohu]] pao]# [$_{VP}$[$_{NP}$ paper [tiger]] runs]#	[zhi-[lao-hu]] pao 3 3 3 3 (3 s s 3) p369	1) p $m_{n.}$ =>(**√lao√hu)n.** (3 3) (s 3) 2) (p $m_{n.}$ M n.) =>(**√zhi√laohu)n.** **(√pao)v.** (3 s 3 3) (3 s s 3)

14.	[$_{VP}$ kao # [$_{NP}$ xiao[ruge]]] [$_{VP}$ roasted # [$_{NP}$ small[squads]]]	"roast small squabs" kao # [xiao ru-ge] 3 3 3 1 (3 s 3 1) p387	(M (p_c) $m_{n.}$)
(p $m_{n.}$) shows the concatenation process inside the compounding structure domain of $A_{N1}[BC]_{N2}$ while M shows the regular M structure above the classical word domain, which is the verb of VP structure here. Actually we will show that there is no need to identify the boundary of the word domain later.			

To figure out the concatenation results of lexical domain in $A_{N1}[BC]_{N2}$ and $A_{Adj}[BC]_N$, it is easier to illustrate **14** in table 5 above as an example with the syntactic tree:

(14) Syntactic tree for $A_{Adj}[BC]_N$—$xiao_{Adj}[ruge]_N$ ($small_{Adj}[squads]_N$)



As we can see from the tree above, the head movements are: the first head movement from √XIAO to a°, the second incorporation movement from aP to √RU-GE, and the final head movement from √ru-ge to n°. aP is a independent structure where √XIAO belongs to phrasal category—P structure, while √RU-GE belongs to non-phrasal category—M structure. None of

the roots obtain c-command relation: (p m $_{n.}$). Together with the verb from VP above the lexical domain, (M # (p_c) $m_{n.}$) is concatenated. This time (p..m) wins.

6.2. Residual problem from data at the sentence level about the competition between two types of concatenation process

The competition between M...P c. algorithm (Head-left Concatenation) and P...M n. algorithm (Phrase-left Concatenation) appears again here below the classical word domain. In a different way, this time P...M n. algorithm (Phrase-left Concatenation) outweighs M...P c. algorithm (Head-left Concatenation). We can temporarily conclude: the structure occupying the lower position in syntactic trees outweighs. The structures in the lower position of syntax tree have priority to construct legitimate concatenation domains in the first place, then the overlapping Morphological words can reconstruct another legitimate concatenation domain for another time with the competitors from the higher position of syntax tree.

Spell-out phases are shown not to be direct boundaries to sandhi domain distribution, which means that they cannot directly block the tone sandhi domain construction. However, in some circumstances, materials spelled out in the early phases have priority when there is a chance to form a legitimate sandhi domain without interference, even with competitors from later phases in some situations. How this characteristic can be related to spell-out phase effects is not clearly known yet.

We look again at the unsolved problems in the former table from **1b1**,**1b2** and **1b3**, they seem to be potentially unified in this lower-syntax-tree-position priority in competition between (M...P c.) algorithm and (P...M n.) algorithm, which is a good outcome:

Between 1b1 and 1b2, the controversial outcomes can be unified with (P (M n.) P c.), with M P c. takes priority since it occupies the deeper position in the syntax tree or in some circumstances the structure belonging to a former cyclic phase at the sentence level.

1b3 can also be explained if we add the incorporated structure of the lexical domain into consideration. The whole structure will be P M n. # p $_{c.}$ m $_{n.}$ structure without the necessary competition between (P (M n.) P c.).

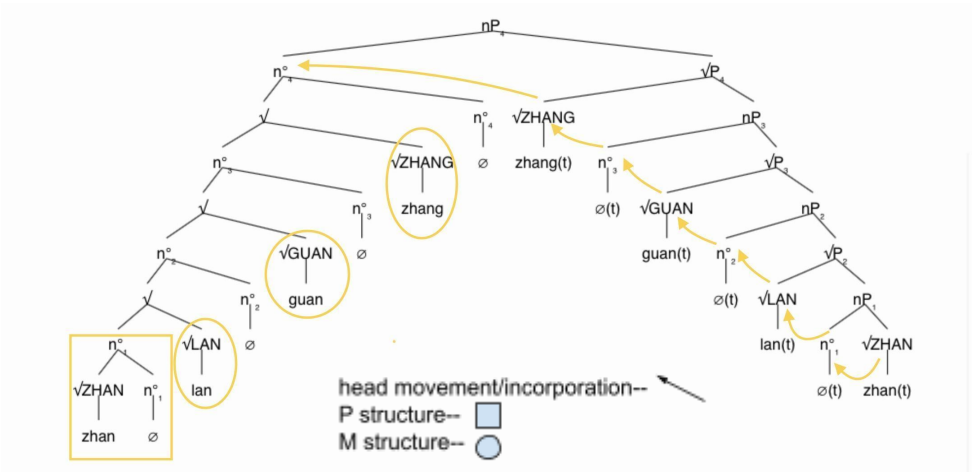
6.3. Four-morpheme compounding structures

Then we turn to the four-morpheme compounding structures, with examples also from Chen (2009)—[[AB]$_{N1}$C$_{N2}$]D$_{N3}$ and A$_{Adj1}$[B $_{Adj2}$[CD] $_{N1}$].

6.3.1. Type I : [[[AB]$_{N1}$C$_{N2}$]D$_{N3}$]

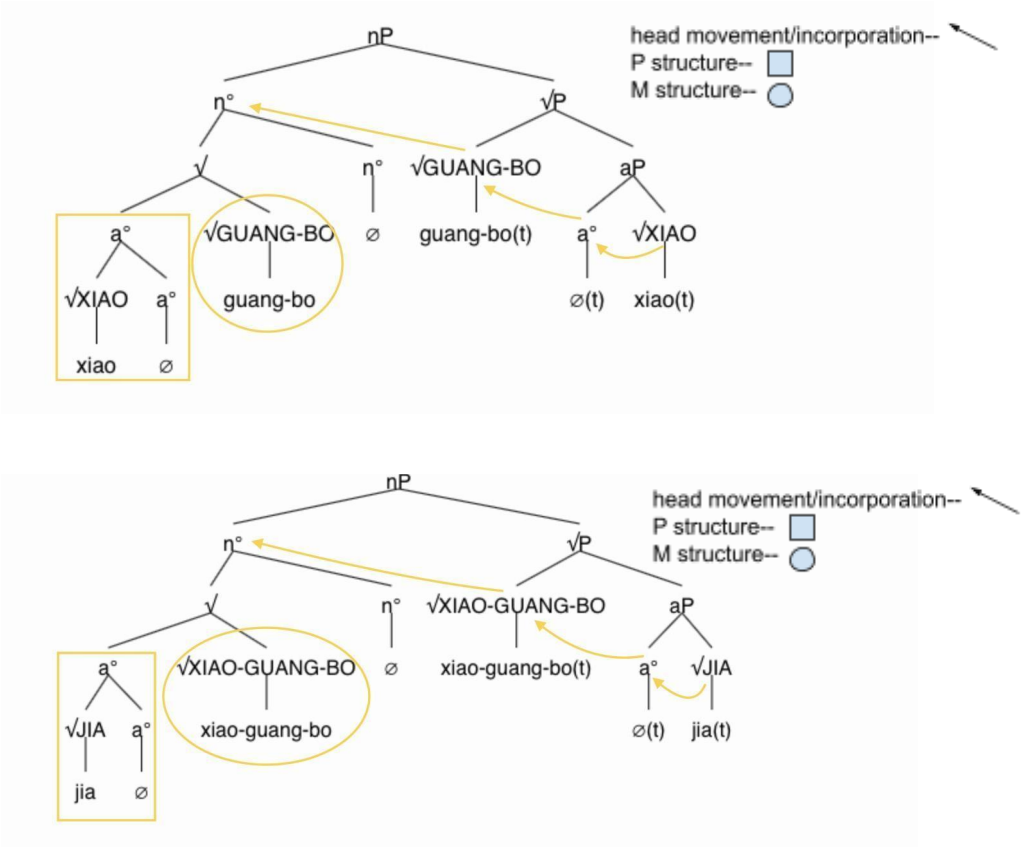
	Consecutive pattern	Separate pattern
Type I : [[[AB]$_{N1}$C$_{N2}$]D$_{N3}$] [[[zhan-lan]$_{N1}$guan$_{N2}$]zhang$_{N3}$] [[[exhibition]$_{N1}$hall$_{N2}$]director$_{N3}$] “director of exhibition hall”	A	B

The stress assignment in Shanghai Chinese has ruled out B type here, so we are only going to check A type. A type is a legitimate result: with continuous multiple incorporation movements, nP$_1$ is an independent structure where √ZHAN belongs to phrasal category—P structure, while √LAN, √GUAN, √ZHANG belong to non-phrasal category—M structure. There is no c-command relation between any two out of these roots: (p m $_{n.}$), producing the right concatenation result: p m$_{n.}$ m$_{n.}$ m$_{n.}$

A.	

6.3.2. Type III : [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]]

	consecutive pattern	separate pattern
Type III: [A_{Adj1} [B_{Adj2}[CD] $_{N1}$]] [jia_{Adj1} [$xiao_{Adj2}$[guang-bo] $_{N1}$]] [$fake_{Adj1}$ [$small_{Adj2}$[rumor] $_{N1}$]] "false rumor"	E	F1 F2

F1.	

The stress assignment in Shanghai Chinese has ruled out E and F2 type here, so we are going to check F1 type only:

F1 type is a legitimate result with separate steps of incorporation movements. In the first step, aP is an independent structure where √XIAO belongs to the phrasal category—P structure, while √GUANG-BO belongs to the non-phrasal category—M structure. None of the roots obtain c-command relation: (p $m_{n.}$), producing the concatenation result: p $m_{n.}$ In the second step, aP is an independent structure where √jia belongs to the phrasal category—P structure, while √XIAO-GUANG-BO belongs to the non-phrasal category—M structure. There is no c-command relation between any two out of these roots: (p $m_{n.}$), producing the concatenation result: p $m_{n.}$

Because of the separate steps of incorporation movements, tone sandhi rule also applies multiple times, promising the priority of tone sandhi between √XIAO and √GUANGBO, eliminating the potential tone sandhi between √JIA and √XIAO.

We conclude the concatenation results of four-root compounds in the table below:

Table 6	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
15.	[[$_{NP}$ [zhan-lan] guan] zhang]# [[$_{NP}$ [exhibition] hall] director]# "director of exhibition hall"	exhibit hall director [[zhan-lan]-guan]-zhang (3 3 3 3) (s s s 3)	p $m_{n.}$ $m_{n.}$ $m_{n.}$ #
16.	[$_{NP}$ jia [$_{NP}$ xiao[guang-bo]]] [$_{NP}$ fake [$_{NP}$ small [radio]]] "false rumor"	false small broadcast jia-[xiao-[guang-bo]] (3 3 3 1) (3 s 3 1)	1) p $m_{n.}$ =>(**√xiao √guangbo)n.** (3 3 1) (s 3 1) 2) p $m_{n.}$ =>(**√jia √xiaoguangbo)n.** (3 s 3 1)

6.4. More evidences in two-morpheme structure

We will rewrite the two-morpheme structure in Table 4 with incorporated compounding structure:

Table 4 adapted	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
1a.	[$_{VP}$ xie [$_{NP}$ xiaoshuo]] [$_{VP}$ write [$_{NP}$ novel]]	write novel xie [xiao-shuo] 3 3 1 (s 3 1)	(M (p $_{c.}$) m $_{n.}$)
1b.	1b1 [$_{CP}$ gou [$_{VP}$ yao wo]] [$_{CP}$ dog [$_{VP}$ bite me]] 1b2 [$_{CP}$ wo [$_{VP}$ xie [$_{NP}$ shu]] [$_{CP}$ I [$_{VP}$ write [$_{NP}$ book]] 1b3 [$_{CP}$ gou [$_{VP}$ yao [$_{NP}$ nü-ren]]] [$_{CP}$ dog [$_{VP}$ bite [$_{NP}$ woman]]]	dog bite me gou [yao wo] (3 (3 3)) (s 3) TS (3 s 3) TS not applicable = reading (i) --- ((3 3) 3) Alternative Foot Formation (s 3) TS (s s 3) TS = reading (j) p418	1b1 and 1b2 (P (M n.) P c.) 1b3 P M n. # p $_{c.}$ m $_{n.}$

		I write book wo [xie shu] 3 3 1 (s 3 1) p377 “the dog bit the woman” [$_{CP}$ gou [$_{VP}$ yao [$_{NP}$ nü-ren]]] (3 3 3 2) (2 3) (3 2) p411	
2.	[$_{VP}$ xiang [$_{CP}$ pro mai gudong #]] [$_{VP}$ want [$_{CP}$ pro buy antiques #]] “(null subject) want to buy antiques ” [$_{VP}$ xiang [$_{CP}$ pro xie xiaoshuo #]] [$_{VP}$ want [$_{CP}$ pro write novel #]] “(null subject) want to write novel” (in Chinese there is no “to” after “want” to occupy the T position in TP.)	want buy antique “wants to buy antiques” xiang [mai [gu-dong]] 3 3 3 3 (2 3) (2 3) want write novel xiang # [xie # xiao-shuo] (3 3 3 1) (s 3) (3 1) Phrasal MRU, TS i = ok (s s) (3 1) Cross-MRU TS optional j = ok p386	(M M c.(p $_{c.}$) m$_{n.}$)

3.	[$_{CP}$ wo [$_{VP}$ xiang # [$_{CP}$ pro mai gupiao #]]] [$_{CP}$ I [$_{VP}$ want # [$_{CP}$ pro buy stocks #]]]		P M n. # (M c (p $_{c}$) m $_{n}$)
5.	[$_{VP}$ mai [$_{NP}$ [$_{PP}$ gei ni # [$_{NP}$ gupiao]]]] [$_{VP}$ buy [$_{NP}$ [$_{PP}$ for you # [$_{NP}$ stocks]]]] "buy you stocks" (in Chinese "for" here is necessary when indirect object is inverted with direct object.)		M M c. P c.# p $_{n.}$ m $_{n}$
6.	[$_{VP}$ mai [$_{NP}$ gupiao # [$_{PP}$ gei ni #]]] [$_{VP}$ buy [$_{NP}$ stocks # [$_{PP}$ for you #]]]		(M (p $_{c.}$)m $_{n}$) # M n. P c.
7.	[$_{CP}$ wo $_{VP}$[xiang # [$_{CP}$ PRO [$_{VP}$ mai [$_{NP}$ gupiao#[$_{PP}$ gei ni#]]]]]] [$_{CP}$ I $_{VP}$[want # [$_{CP}$ PRO [$_{VP}$ buy [$_{NP}$ stocks # [$_{PP}$ for you#]]]]]]		P M n. # (M (p $_{c.}$)m $_{n}$) # M n. P c.#
8.	[$_{baP}$ ba [$_{VP}$ [$_{NP}$ bi] # gei ni #]]] [$_{baP}$ ba [$_{VP}$ [$_{NP}$ pen] # give you #]]] "give you the pen" (ba structure: ba +direct object +verb+indirect object =verb + direct and indirect objects) Bap takes vp as compliment	Xiaomei BA pen give Xiaoming "Xiaomei gave the pen to Xiaoming" [xiao-mei] [[ba bi] [gei xiao-ming]] 3 3 3 3 3 3 2 (2 3) (2 3) (2 3 2) p407	M P c.# (M (p $_{c.}$)m $_{n}$)

12	[$_{CP}$ gou [$_{VP1}$ chao **xing** # [$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] [$_{CP}$ dog [$_{VP1}$ barks **wakes up** #[$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] “the dog barks so that Xiaomei wakes up” Verb incorporation Structure of verb + resultative verb	dog noisy-wake up “the dog woke up Xiaoming by making noise” gou $_{VP}$[$_{V}$[chao-xing] $_{NP}$[xiao-ming]] 3 3 3 3 2 (s 3) (3 2) Lexical MRU, TS (3 s 3) (3 2) Postlexical MRU, TS not applicable (optional between MRUs) p391	(P (m $_{n.}$)p $_{c.}$) # p $_{c.}$ m $_{n.}$

6.5. Exceptions in Chen (2009) can be unified in the current framework.

We will now talk about the exceptional situations in Chen(2009) in the lexical and postlexical two-pass MRU (minimal rhythm unit) framework, which are shown to be well predicted in the current framework.

Table 7	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
Phonological word (Cliticization)			
17.	[$_{VP}$[$_{PP}$[$_{NP}$zhong-tong-fu]li] you] [$_{VP}$[$_{PP}$[$_{NP}$president palace]inside] have]	president palace inside have [[zong-tong]-fu]-li # you (zong-tong-fu = li) 3 3 3 3 3	p m$_n$ m$_n$ M$_n$ M$_n$

	"(as for famous paintings), there are (quite a few) inside the presidential palace"	(s 3) Lexical MRU, TS (s s 3) Lexical MRU, TS (s s s 3) Lexical MRU, TS (s s s s 3) Phrasal MRU, TS p398	
18.	mao $_{VP}$[$_{PP}$[bi gou] xiao] cat $_{VP}$[$_{PP}$[than dog] small] "the cat is smaller than the dog"	cat than dog small mao $_{VP}$[$_{PP}$[bi gou] xiao] mao = bi $_{VP}$[gou xiao] Cliticization 1 3 3 3 (1 3) Lexical MRU (1 3) (s 3) Phrasal MRU, TS p399	P M n. # P n. M n.
19.	[$_{VP}$ mai [$_{CIP}$ dian [jiu]]] [$_{VP}$ buy [$_{CIP}$ some [wine]]] "buy some wine" CIP: Classifier phrase	buy some wine mai [$_{NP}$ dian jiu] (mai = dian) jiu Cliticization 3 3 3 (s 3) Lexical MRU, TS	M M c P c

		(s s 3) Phrasal MRU, TS p401	
20.	[NP fang-dong][NegP bu[VP yun-xu [IP[NP wo] [VP PRO yang xiao-mao]]]] [NP land lord][NegP not[VP allow [IP[NP me] [VP PRO keep kitten]]]] "the land lord doesn't allow me to keep kittens"	[NP fang-dong][NegP bu [VP yun-xu [IP[NP wo] [VP PRO yang xiao-mao]]]] (yun-xu = wo) Cliticization 2 1 4 3 3 3 3 3 1 base tone (2 1) (s 3) (3 1) Lexical MRU, TS (s s 3) Lexical MRU, TS ---------- -(2 1) (4 s s 3) (s 3 1) Phrasal MRU, TS p402	p m n # (M n (p c)m n M n) # (M c (p c)m n)
Syntactic word (modifier+noun /complex predicate)			
21.	[VP kao # [NP xiao[ruge]]] [VP roasted # [NP small[squads]]]	"roast small squabs" kao # [xiao ru-ge] 3 3 3 1 (3 s 3 1) p387	(M # (p c) m n.)
22.	[CP gou [VP1 chao **xing** # [VP2 [NP Xiaomei ___]]]]]	dog noisy-wake up "the dog woke up Xiaoming by making noise"	(P (m n.) p c.) # p c. m n.

	[$_{CP}$ dog [$_{VP1}$ barks **wakes up** #[$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] “the dog barks so that Xiaomei wakes up” Verb incorporation Structure of verb + resultative verb	gou $_{VP}$[$_{V}$[chao-xing] $_{NP}$[xiao-ming]] 3 3 3 3 2 (s 3) (3 2) Lexical MRU, TS (3 s 3) (3 2) Postlexical MRU, TS not applicable (optional between MRUs) p391	

The exceptional situations of MRU (minimal rhythm unit) framework in Chen (2009) are summarized below:

a. 17 to **20** are exceptional situations about Phonological words (Cliticization). **17** and **18** belong to the subtype of prepositions. In **17**, the clitic of preposition “li (inside)” are determined by Chen (2009) to be combined with “zong tong fu (president palace)” to be a single Phonological word, where the lexical MRU (minimal rhythm unit) algorithm applies, instead of processing the preposition “li (inside)” in postlexical MRU (minimal rhythm unit) algorithm. In **18**, the clitic of preposition “bi (compare)” is determined by Chen (2009) to be combined with the matrix subject “mao (cat)” to be a single Phonological word, where the lexical MRU (minimal rhythm unit) algorithm applies as well. The clitic in **19** belongs to the subtype of classifiers: the clitic of classifier “dian (some)” is determined by Chen (2009) to be combined with the verb “mai (buy)” to be the same Phonological word, instead of the object “jiu (wine)”, where the lexical MRU (minimal rhythm unit) algorithm takes the priority to apply firstly in “mai=dian (buy some)”. The clitic in **20** belongs to the subtype of object pronouns: the clitic of object pronoun “wo (me)” is determined by Chen (2009) to be combined with the matrix verb “yun-xu (allow)” to be the same Phonological word, instead of the verb in the subordinate clause “yang (keep), where the the lexical MRU (minimal rhythm unit) algorithm applies firstly between “yun-xu=wo (allow me keep)”.

There seems to be some problems about the approaches to these exceptional situations in Chen (2009). In **17, 18**, **19** and **20**, whether the prepositions, classifiers and the object pronouns can be viewed as integrated lexical domain with the former lexical domain is not clear. If we don't misunderstand, in Chen (2009) the lexical tone sandhi domain is constructed in the classical lexicon following the classical cyclicity in the word formation. In this case, whether these Phonological words (prepositions, classifiers or object pronouns) can be assumed to combine with former lexical words to be an integrated lexical domain is not clear. In other words, if we include these exceptional situations to be considered as integrated lexical domains under Chen's framework, the lexical boundary will no longer be the classical lexical domain in the literature. The special lexical boundary may lack a redefinition. Meanwhile, in **20**, The object pronoun "wo (me)" and the matrix verb "yun-xu (allow)" share a closer syntactic relation: the object pronoun "wo (me)" is under movement from DP of VP in subordinate clause to DP of IP, which can potentially have a closer relation with the matrix verb "yun-xu (allow)" instead of the verb in the subordinate clause "yang (keep).

b. In **21** and **22**, exceptional situations about Syntactic words are discussed. **21** is modifier+noun structure and **22** is complex predicate. In **21**, Chen (2009) advocates that modifier+noun structure belongs to the word domain similar to relevant discussions in the literature. In this way, modifier+noun structure belongs to a single syntactic word, where the lexical MRU (minimal rhythm unit) algorithm applies, instead of processing the modifier (adjective) "xiao (little)" in postlexical MRU (minimal rhythm unit) algorithm. In **22**, Chen (2009) advocates that (verb+resultative verb) double-verb structure belongs to the word domain, forming a single syntactic word where the lexical MRU (minimal rhythm unit) algorithm takes the priority to apply between "chao (bark)" and "xing (wake up)".

In our current framework, it seems not necessary to reach out to the Phonological word and Syntactic word in Chinese languages, part of which also has controversial discussion in the literature.

In **17**, NP takes the position of specifier of PP, which also takes the position of specifier of VP. NP is a three-root compounding structure, with concatenation results of ($p\ m_n\ m_n$). P of PP belongs to the non-phrasal category. V of VP belongs to the non-phrasal category. None of the Morphological words belongs to c-command relation, so the final concatenation result is ($p\ m_n\ m_n\ M_n\ M_n$).

In **18**, PP also takes the position of specifier of VP. NP “gou (dog)” takes the position of complement of PP. NP “mao (cat)” takes the position of specifier of a higher phrase than VP, e.g. IP or CP. N of NP “mao (cat)” belongs to the phrasal category. P of PP belongs to the non-phrasal category. N of NP “gou (dog)” belongs to the phrasal category. V of VP belongs to the non-phrasal category. None of the Morphological words belongs to c-command relation, so the final concatenation result is (P M n. # P n. M n.).

In **19**, CIP takes the position of complement of VP. NP takes the position of complement of CIP. V of VP belongs to the non-phrasal category. CI of CIP belongs to the non-phrasal category. N of VP belongs to the phrasal category. All the Morphological words belong to c-command relation, so the final concatenation result is (M M c P c.).

20 is carefully illustrated and discussed in section 7.2. **21** is explained in section 6.1. and **22** is explained in section 5.2. in example 12.

6.6. Summary of section 6

In section 6, we realize the mutual tests with the multiple-root incorporated structure from section 3 and the framework of T3 tone sandhi domain construction as Concatenation rule in Mandarin Chinese from section 5. In section 6.1, we test the three-morpheme compounding structures: $A_{N1}[BC]_{N2}$ and $A_{Adj}[BC]_{N.}$,adding the multiple-root incorporated structure into the results of the concatenation process and producing compatible outcomes. In part 6.2, we solve the residual problem from the data in sentence level about the competition between two types of concatenation process-Head-left Concatenation and Phrase-left Concatenation. It shows that the structures in the lower position of syntax tree have priority to construct legitimate concatenation domains in the first place, then the overlapping morphological word can reconstruct another legitimate concatenation domain again with the competitors from the higher position of syntax tree. In part 6.3, we test the four-morpheme compounding structures: $[[AB]_{N1}C_{N2}]D_{N3}$ and $A_{Adj1}[B_{Adj2}[CD]_{N1}]$, which is shown to be compatible with the multiple-root incorporated structure in section 3. In part 6.4, we rewrite the data with two-morpheme structures in Table 4 to be compatible with the current multiple-root incorporated structure. In part 6.5, we list Table 6 of the main exceptions in the classical two-step lexical-postlexical MRU-based analysis from Chen (2009) that can be unified in the current framework integrated with multiple-root incorporated structure, showing promising results.

7. General discussion

7.1. Four-morpheme compounding structures in Chinese

One important outcome from the current thesis is that we propose and test the relevant syntactic incorporated structures for Chinese compound words. We repeat table 1 of the four-morpheme compound structure in Chinese in section 4.5. Please refer to the corresponding syntactic trees of detailed multiple-root incorporated structure back in section 3. The letters marked with red color are the structures with successful incorporation patterns.

Table 1	Consecutive pattern	Separate pattern
Type I : $[[[AB]_{N1}C_{N2}]D_{N3}]$	A	B
Type II : $[[A_{Adj1}[BC]_{N1}]D_{N2}]$	C	D1 D2
Type III: $[A_{Adj1}\,[B_{Adj2}[CD]_{N1}]]$	E	F1 F2

7.2. Concatenation algorithm review

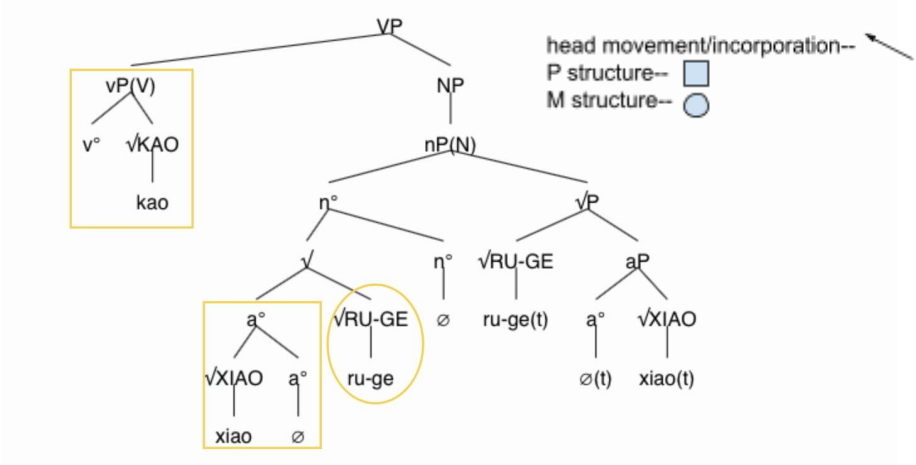
Before we summarize the final Concatenation rule for the domain construction of T3 Tone sandhi in Chinese Mandarin. An important question needs to be addressed in this conclusion section.

7.2.1. Opaque monomorphemic structures in the Concatenation algorithm?

In the current framework —Distributed Morphology, the word domain is extended to be a syntactic structure in the tree with root incorporating to its own category head, e.g., a° v° n°. In this way the lowest position of each phrase in the tree has changed to be the root position.

If we want to stick to the basic concept of Distributed Morphology, we have to consider the incorporated movement structure for all the monomorphemic structures instead of only compounding structures with equal or more than two morphemes. In this case, the morphological relation of relative morphological characteristics which were discussed above in the separate tables need to be reconstructed again. But we are not going to do this. We will explain why we think it is not necessary. Please take a look at (15) from table 5 below.

(15) Syntactic tree for [kao$_{v}$[xiao$_{Adj}$[ruge]$_{N}$]] ([roast$_{v}$[xiao$_{Adj}$[small$_{Adj}$[squads]$_{N}$]])



[$_{VP}$ kao # [$_{NP}$ xiao[ruge]]]

[$_{VP}$ roast # [$_{NP}$ small[squads]]]

In the example (12), the monomorphemic verb √KAO is also illustrated in the (category head+root) structure, corresponding to the basic concept from syntactic word formation in Distributed Morphology. In this modified structure with the current monomorphemic structure (category head+ root structure), problems occur when the M and P structures can no longer be able to represent the morphological relations. In this way, the algorithm no longer describes the morphological relation between N of NP or V of VP, but the relation between √root positions below N(nP) or V(vP), e.g., in the example above, the algorithm should start from the √KAO root position instead of V(vP) node as we regulated at the sentence level discussion in Table 4. This will change the concatenation results: √KAO belongs to **(p)** morphological locality while V(vP) node belongs to **(M)** morphological locality. Since every monomorphemic structure will naturally take **(p)** locality for the √root position (without any

head movement or incorporation movement to construct “m” locality), the simple structure (p…m) cannot illustrate the morphological relation here.

We have one possible way to solve this:

A complex (P…M) plus (p…m) algorithm can be reconstructed: “M” and “P” are used to describe the morphological locality structure between classical syntactic heads, e.g., N of NP or V of VP, and “m” and “p” are used to describe the morphological locality structure between√root positions below N(nP) or V(vP) heads (in the classical view of word domain). We take a few examples to develop this possibility (the examples are taken from the tables before, however they are reordered with new numbers for better readability):

Table 8	***Example sentences***	***Tone sandhi results in Chen (2009) with page numbers***	***Concatenation results marked with symbols denominated in Table 3***
22.	[$_{VP}$ kao # [$_{NP}$ xiao[ruge]]] [$_{VP}$ roast # [$_{NP}$ small[squads]]]	“roast small squabs” kao # [xiao ru-ge] 3 3 3 1 (3 s 3 1) p387	(Mp (Pp)$_c$ Pm)$_n$ —> p(M P)$_c$ P(p m)$_n$
23.	[$_{VP}$[$_{NP}$ zhi [laohu]] pao]# [$_{VP}$[$_{NP}$ paper [tiger]] runs]#	[zhi-[lao-hu]] pao 3 3 3 3 (3 s s 3) p369	Pp Pm$_n$ Pp$_n$ —> P(p m)$_n$ ~~P(m p)$_n$~~
24.	[$_{VP}$ xie [$_{NP}$ xiaoshuo]] [$_{VP}$ write [$_{NP}$ novel]]	write novel xie [xiao-shuo] 3 3 1 (s 3 1) p377	(Mp (Pp)$_c$ Pm)$_n$ —> p(M P)$_c$ P(p m)$_n$

25.	[$_{VP}$[$_{PP}$[$_{NP}$zong-tong-fu]li] you] [$_{VP}$[$_{PP}$[$_{NP}$president palace]inside] have] "(as for famous paintings), there are (quite a few) inside the presidential palace"	president palace inside have [[zong-tong]-fu]-li # you 3 3 3 3 3 (s 3) Lexical MRU, TS (s s 3) Lexical MRU, TS (s s s 3) Lexical MRU, TS (s s s s 3) Phrasal MRU, TS = (73-i) p398	$Pp_n\ Pm_n\ Pm_n\ Mp_n\ Mp_n$ —> $P\ (p\ m\ m)_n\ Mp_n\ Mp_n$?
26.	[$_{CP}$ gou [$_{VP1}$ chao **xing** # [$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] [$_{CP}$ dog [$_{VP1}$ barks **wakes up** #[$_{VP2}$ [$_{NP}$ Xiaomei ___]]]]] "the dog barks so that Xiaomei wakes up" Verb incorporation Structure of verb + resultative verb	dog noisy-wake up "the dog woke up Xiaoming by making noise" gou $_{VP}$[$_{V}$[chao-xing] $_{NP}$[xiao-ming]] 3 3 3 3 2 (s 3) (3 2) Lexical MRU, TS (3 s 3) (3 2) Postlexical MRU, TS not applicable (optional between MRUs) p391	Pp # $Mm_n\ Mp_c$ # $Pp_c\ Pm_n$ —> Pp # $M(p\ m)_n$ # $P(p\ m)_n$
27.	[$_{DP}$ nei[$_{CIP}$ zhong [$_{NP}$ jiu]]] [$_{VP}$ you-hai] [$_{DP}$ that[$_{CIP}$ kind [$_{NP}$ wine]]] [$_{VP}$ harmful]	[$_{DP}$ nei[$_{CIP}$ zhong [$_{NP}$ jiu]]] [$_{VP}$ you-hai] 4 3 3 3 4	Pp Mp_n Mm_n# Pp_c Pm_n $p(P\ M)_n\ M(p\ m)_n$# $P(p\ m)_n$

	"that kind of wine is harmful"	(4 s 3) (3 4) p372	

Most structures are compatible with the double M/P plus m/p system, except the example 25, which may be a coincidence. With this new system, we can reformulate the VP structures from 22 to 24 perfectly. As we can see, all the $p(M\ P)_c$ / $P(p\ m)_n$ /$M(p\ m)_n$ …...can be the legitimate concatenation results for the complex tone sandhi algorithm while $P(m\ p)_n$ leads to an unsuccessful result. Logically, we can attempt to list all the possible legitimate concatenation results for two-morpheme (root) structure:

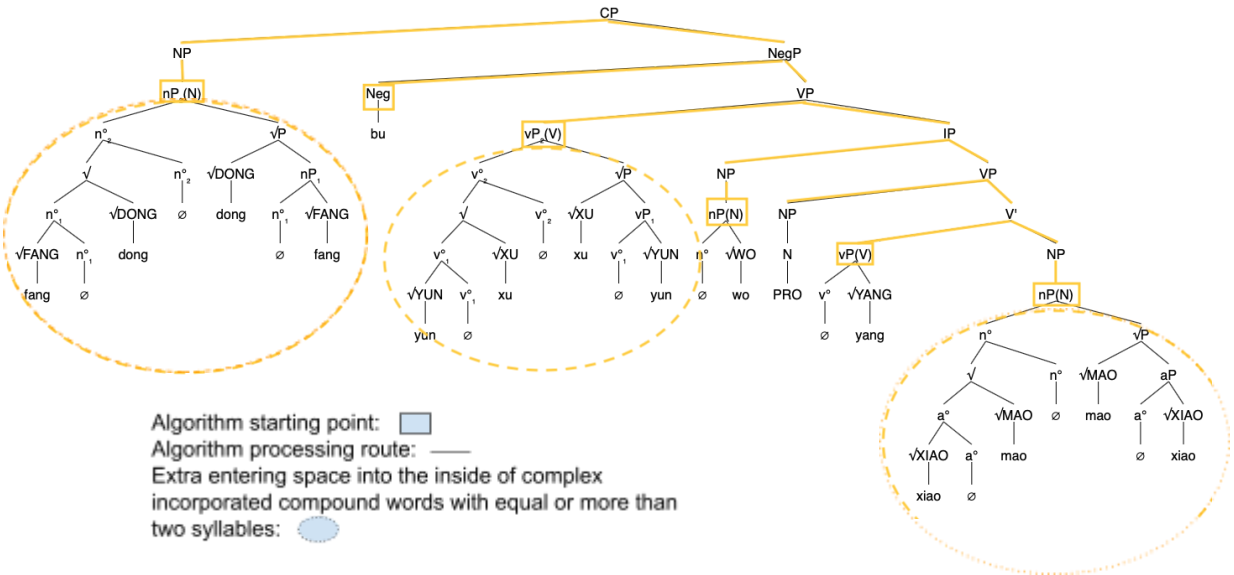
Table 9	$(p\ m)_n$/$(P\ M)_n$	$(m\ p)_c$ /$(M\ P)_c$
M/P	$P(p\ m)_n$ $M(p\ m)_n$	$P(m\ p)_c$ $M(m\ p)_c$
m/p	$p(P\ M)_n$ $m(P\ M)_n$	$p(M\ P)_c$ $m(M\ P)_c$

This table 9 above shows the 8 possible legitimate concatenation results only for two-morpheme (root) structure, with 4 of them having been tested from limited examples emphasized in red. I cannot deny the possibility that these 8 concatenation results can be the basic functions of the tone sandhi algorithm to process sandhi domain construction. However, this double system shows no obvious better performance than our previous analysis: **a.** We start the calculation from the √**root position** below N(nP) or V(vP) **only** in the compounding structure with 2 or above morphemes. **b.** We start the calculation from the node N(nP) or V(vP) for the large number of monomorphemic structures. The 8 possible legitimate concatenation results greatly increase the processing complexity than the only 2 $(p\ m)_n$/$(P\ M)_n$ and $(m\ p)_c$ /$(M\ P)_c$ (We didn't distinguish the difference between the concatenation results marked in uppercase and lowercase letters in previous analysis, which is just for registration convenience). In this double system, to distinguish the position between N(nP) or

V(vP) and the root position below N(nP) or V(vP), we have to make clear the strict difference, e.g., between P(p m)$_n$ and p(P M)$_n$. It seems to be not necessary to increase complexity by resetting the starting point of the algorithm to the root position for all the monomorphemic structures. Since all the monomorphemic structures naturally form the united P structure, there will be much wasted work for non-changeable united P structure, e.g., for many phrases only with monomorphemic structures. It seems to be not necessary to do this updated complex double system to process this kind of phrase. For language processing, saving unnecessary energy in every possible way seems to be the right and better choice since there are so many other factors to process.

We take a look at example **20** in Table 7 to illustrate this process with shifty algorithm starting points in (16) (**a. root position** below N(nP) or V(vP) **only** in the compounding structure with 2 or above morphemes. **b.** the node N(nP) or V(vP) for the large number of monomorphemic structures):

(16) Syntactic tree for example **20** in Table 7



p m $_n$ # (M $_n$ (p $_c$)m $_n$ M $_n$)# (M $_c$ (p $_c$)m $_n$)

[$_{NP}$ fang-dong][$_{NegP}$ bu[$_{VP}$ yun-xu [$_{IP}$[$_{NP}$ wo] [$_{VP\ PRO}$ yang xiao-mao]]]]

[$_{NP}$ landlord][$_{NegP}$ not[$_{VP}$ allow [$_{IP}$[$_{NP}$ me] [$_{VP\ PRO}$ keep kitten]]]]

“the landlord doesn’t allow me to keep kittens”

We use the syntax tree of example **20** in Table 7 to illustrate the possible algorithm in the application process of T3 tone sandhi in Mandarin Chinese. As we can see in the tree above, to form the morphological concatenation chain, the regular algorithm starting point is from head position of each phrase in the syntax tree, e.g., V from VP, N from NP, which remains to be the lowest position of each phrase in the standard form of classical syntax domain. From the discussion above, for both monomorphemic structures and compounding structures with equal and more than two morphemes, setting the united algorithm starting point in root position seems to complicate the application process by adding extra basic functions in calculation and showing no better accuracy, convenience or economy in the application process. As a result, we turn to the basic form—the algorithm starting point moves from head position of each phrase in the classical syntactic domain (e.g. V from VP, N from NP) to the root position, **only** in compounding structures with equal or more than two morphemes, **not** in the monomorphemic structures. In other words, the algorithm processing route goes smoothly between the head positions of each phrase in the classical syntactic domain (e.g., V from VP, N from NP), **only** diving into the compounding structures with equal or more than two morphemes under syntactic word structure, **not** in the monomorphemic structures. In this example, we have the concatenation results—p m $_n$ # (M $_n$ (p $_c$)m $_n$ M $_n$)# (M $_c$ (p $_c$)m $_n$), with the lowercase letters showing the algorithm diving into the compounding structures, and the uppercase letters showing the regular basic algorithm starting point in the classical syntactic head position in the tree. The result shows that there are two definite tone sandhi boundaries dividing three tone sandhi domains, and two of which contain the competition between (M...P c.) algorithm (Head-left Concatenation) and (P...M n.) algorithm (Phrase-left Concatenation), which was discussed in Section 6.2. In this phrase, the algorithm of tone sandhi domain dives three times into the compounding structure—the NP subject of matrix clause, the VP predicate of matrix clause and the NP object of the subordinate clause, showing good performance and function in this way of calculation.

Under the current framework with phonological rule to be specific of T3 tone sandhi in Mandarin Chinese, I think that we can temporarily conclude: the monomorphemic structures are opaque in the application process (the algorithm of concatenation rule) of T3 tone sandhi in Mandarin Chinese, and it is shown to be a product or outcome from certain phonological rules’ processing economy and convenience: **a).** Not diving into the large amount of

monomorphemic structures greatly saves the processing energy for the algorithm. **b)**. Such opaque monomorphemic spaces are indispensable for the necessary benefits of high economy, convenience and efficiency.

As a reminder, this speculation is not contradicted by the fact that monomorphemic structures still adhere to the basic syntactic word domain—root and "category head" structure, but these monomorphemic structures are not the space in the syntactic tree where the algorithm of domain construction for tone sandhi will dive, for the processing efficiency and economy. I would rather refer to this processing strategy as a superpower ability of language.

7.2.2. Are these opaque monomorphemic structures related to the classical dispute about "word"?

Based on the discussion in section 7.2.1, under current approach, we may have a potential motivation to explain why we have the **"word" concept**: for some certain phonological rules, the monomorphemic structures are structurally "left over". Under current syntactic word formation, some certain phonological rules create the distinctive opaque spaces, different from any other spaces in the tree, which could be the corresponding **"word" space.** In this case, this "word" concept is not a true grammatical identity, but a product or outcome of certain phonological rules' processing economy and convenience.

Between lexicon-based word formation theory and new word formation theories such as syntactic word formation theory, there is always a dispute about whether we have "word" or not and what is the definition of "word" that we are talking about. As we use the syntactic word formation theory all the way in the current work, these opaque monomorphemic structures seem to be more important in the question about whether we have "word" or not. These opaque monomorphemic structures are not grammatically different from any structures or spaces in the syntactic tree—it is shown that a more complex processing algorithm can be chosen to add the monomorphemic structures into calculation by adding more basic functions. This possibility is ruled out due to the extra functions not increasing accuracy but lowering the processing economy of the algorithm. However, they are structurally different from other spaces in the syntactic tree—they are naturally "left-over" opaque structures. In other words, these monomorphemic structures are grammatically similar to, but structurally different from other structures or spaces in the syntactic tree. The grammatical similarity shows the essences in the syntactic word formation theory—there is no grammatical identity

as a “word”, which is constructed under syntactic movements like other larger structures. The structural difference may still align with what lexicon-based approaches argue—“word” identity is still meaningful. In the context of syntactic word formation theory, this meaning of “word” identity may be more structural than grammatical. We may need to talk more about what we are referring to when we talk about the “word”.

Back to the dispute between lexicon-based word formation theory and new word formation theories such as syntactic word formation theory: **a.** The lexicon-based frameworks have been shown not to contribute enough to many phonological phenomena, e.g., phrasal lexical phonological rules. **b.** The syntactic word formation theory seems not to be able to get rid of “word” identity or structure completely yet. They still cannot beat each other down. It seems that we will still get involved in the classical debate for some more time.

7.2.3. Additional information about “opaque monomorphemic structures” in other Concatenation rules in the literature

About monomorphemic structure not being calculated in Concatenation rule under current approach, we can seek for more evidence in the literature.

a. In Pak (2008), the algorithm of Head-left Concatenation is not extended into the inside of monomorphemic structure and the algorithm basic calculation point remains to be in the head position of each phrase in the classical syntactic domain (e.g. V from VP, N from NP). We will first introduce a Concatenation rule—Low-Tone Deletion (LTD) from example (17), then we use example (18) to illustrate the algorithm's basic calculation process. The examples are directly from Pak (2008).

In Pak (2008), “Luganda has a rule of Low-Tone Deletion (LTD) that potentially applies between two HnLn words; when LTD applies, the L on the first word is deleted and a H-plateau is formed between the two words. LTD applies in (17), where a HL verb is followed by a HL object.”

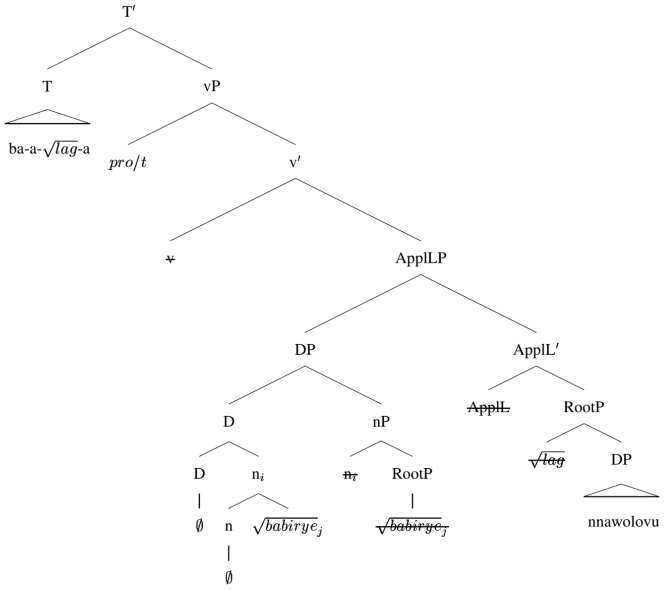
(17) taken from Pak (2008 : 15)

LTD between verb and object:

y-a-fúúmb-à nnawólòvu → (yàf**úú**mb**á** nn**á**w**ó**lòvù)
sbj1-pst-cook-ind 1.chameleon
'S/he cooked a chameleon.'

(18) taken from Pak (2008 : 199)

ba-a-lag-a Babirye nnawolovu
2-pst-show-ind 1.Babirye 1a.chameleon
'They showed Babirye a chameleon.'



In the tree structure above, the algorithm of Head-left Concatenation goes between the head positions of each phrase in the classical syntactic domain (e.g. V from VP, N from NP). When processing D [Babirye] and D [nnawolovu], the algorithm is not diving into nP or RootP to search for the internal structure of these monomorphemic structures. Since the discussion in Pak (2008) is not involved with compounding structures, we could´t know if the Head-left Concatenation algorithm will dive into incorporated compounding structures from the current available data.

However, from Pak (2008) and our analysis above, we know that: **1)**. The original Concatenation algorithm does have the calculation basic point to be the head positions of each phrase in the classical syntactic domain (e.g. V from VP, N from NP), which is consistent with our analysis above to rule out the possibility of resetting the calculation point to be the lower root position in the newly regulated syntactic word domain or forming the

alternative complex double M/P plus m/p system. **2)**. Monomorphemic structures are opaque in the Concatenation algorithm, e.g. in Low-Tone Deletion (LTD), which also form the "left over" space in the tree as what we already discussed in the T3 tone sandhi in Chinese Mandarin.

b. From Prosodic Phonology, there are also approaches to divide the classical phonological word into Single Prosodic Word and Phonological Word Group (Vigário 2010). In Vigário (2010), It is reported that abundant data shows that some combinations of Prosodic words form a "Group", that constitutes the domain of application of specific rules, and which have distinctive properties from Prosodic words.

We use the stress-initial compounds in English from Vigário (2010) to show this approach: At prosodic word level the general rule for stress assignment refers to the right edge of the Prosodic word instead of the left edge. And a certain class of compounds in English shows "a stress pattern that differs from that of φ (Phonological phrase). Whereas φs (Phonological phrase) usually show final prominence, compounds regularly exhibit initial stress."

(19) taken from Nespor (1999b: 138)

Strong-weak pattern in English compounds		
Nominal compounds	*blackbird*	*rádio station*
Adjectival compounds	*cólorblind*	*séasick*
Verbal compounds	*áir condition*	*týpe write*

Vigário (2010) proposes that these facts show that in English there are certain word combinations formed of Prosodic words which do not pattern like other sequences of Prosodic words within φ (Phonological phrase). In this way, these stress-initial compounds are where the "Prosodic words groups are built and stress is assigned on a regular (phonological) basis to the initial Prosodic word of the resulting Prosodic words groups."

We are not going to do the strict mapping from Morphological word under Distributed Morphology with Phonological word under Prosodic Phonology here, while some related work has been done, e.g. Shwayder (2014a,b). There is also some discussion against the possibility or necessity to do the mapping, e.g., Pak (2008) argues that "phonological rules operate directly on the syntactic structure as it happens to exist at that particular point in the PF derivation".

(20) Proposal arguing for phonological rules operating directly on the syntactic structure, taken from Pak (2008 : 20)

"All phonological rules apply *directly* to the syntactic structure as it happens to exist at the given point in PF. Phonological rule domains 'come for free'; there is no need for specially derived prosodic constituents like the Phonological Phrase."

We are not going to extend our discussion about this point since the current T3 tone sandhi in Chinese Mandarin is a non-cyclic phonological rule, which is generally not sensitive to different prosodic domains. However from the parallel theory of Prosodic Phonology, we can also know that there are more approaches to treat classical compounding structures or constructions different from monomorphemic structures as well as phrasal structures. In other words, compounding structures do show distinctive phonological features in some phonological rules.

Overall, we can try to temporarily conclude that:

a. Syntactic incorporated compounding structure is the key to decipher domain construction for T3 tone sandhi in Chinese Mandarin in the framework of Concatenation rule under Distributed Morphology.

   The Concatenation algorithm diving into the syntactic incorporated compounding structure (**not monomorphemic structures**) under current approach is the key to unify the T3 tone sandhi in Chinese Mandarin both below and above classical word domain, ruling out the necessity of multiple rule application processes divided into two steps—lexical level and postlexical level, which is consistent with the economy principle.

b. Compounding structures/constructions are shown to have distinctive phonological features, which are different from monomorphemic structures/constructions as well the phrasal structures/constructions. This was also proposed by the parallel theory—Prosodic Phonology. And this was proved in the current approach, with the Concatenation algorithm only diving into the incorporated compounding structure but not into the monomorphemic structures.
c. Our current approach is the Concatenation rule with direct reference to morphosyntactic information to predict tone sandhi domains in the syntactic word

formation theory under the scope of Distributed Morphology. In this definite approach, monomorphemic structures are "left-over" spaces in the Concatenation algorithm process, which are exempted from some phonological rule application processes. If we take a radical perspective, this may show that monomorphemic structures are the **only** structurally-defined "word" domain under current approach.

7.3. T3 tone sandhi domain construction as Concatenation rule in Mandarin Chinese final checking

With all the discussions in section 6 and 7.2, we will revise the domain construction as Concatenation rule for T3 tone sandhi in Mandarin Chinese in section 5.4. below. The main revisions lie in: **a.** the residual problem about the competition between Head-left Concatenation (illustrated in M...P c.) and Phrase-left Concatenation (illustrated by P...M n.) discussed in section 6.2, which is clarified in point 5) below. **b.** the shifty calculation points of algorithm route discussed in section 7.2, which is clarified in point 2) below.

1). In each sentence, **M** structure from left to right searches for the possibility of Head-left Concatenation (illustrated in M...P c.): identifies pairs of Morphological words X, Y where (i) X is left-adjacent to Y, and (ii) X c-commands Y, and **P** structure from left to right searches for the possibility of Phrase-left Concatenation (illustrated by P...M n.): identifies pairs of morphological words X, Y where (i) X is left-adjacent to Y, and (ii) X does not c-command Y. (**M** and **P** are syntactic structure locality determined by contained Morphological words)

2). The algorithm starting point is the head position of each phrase in the classical syntax domain (e.g. V from VP, N from NP). Only when processing a compounding structure with equal or more than two morphemes, the algorithm dives into the inner space of incorporated compounding structure structure, resetting the starting point to be the root position of the syntactic word domain (root and category head structure). There is no legitimate processing space for the algorithm in the large number of monomorphemic structures.

3). For Concatenation statements continues as A⌢B⌢C…, a sandhi domain boundary is inserted when no further Concatenation statement exists, in other words a sandhi domain boundary is inserted after the ultimate morphological word which can comply with current Concatenation algorithm.

4). Once one sandhi domain is determined, the subsequent new concatenation process is continued from the next Morphological word, repeating the process in 1), until the end of the sentence.

5). In structure (P... (M n.)... P c.), there is competition between (M...P c.) algorithm (Head-left Concatenation) and (P...M n.) algorithm (Phrase-left Concatenation): the structure occupying the lower position in syntactic trees outweighs. Even though spell-out phases are not shown to have direct effects in tone sandhi domain distribution, in some circumstances, materials spelled out in the early phases have priority when there is a chance to form a legitimate tone sandhi domain without interference, even with competitors from later phases.

6). Tone sandhi rule as Chaining operation takes place to establish linear order across all the Morphological words, from left to right within that chain of Morphological words in every sandhi domain.

7). Following Chen (2018), we determine tone sandhi rule in Mandarin Chinese changing from citation form T3 to sandhi form T2: T3—> T2 / ___T3.

Overall, we propose a syntactic multiple-root incorporated structure for Chinese compounding structures adapted from Harley (2006), and a simplified framework of Concatenation rule after linearization of Morphological words applied in domain construction for T3 tone sandhi in Mandarin Chinese determined by specific morphosyntactic information (morphosyntactic locality characteristics and c-command relation) under Distributed Morphology.

Under the current syntactic word formation framework, it shows that the multiple pure N compounding structures (e.g. N-N-N-N) in Chinese are constructed by consecutive multiple incorporation movements forming a consecutive multiple-root incorporated structure. And the multiple A-N compounding structures (e.g. $[[A_{Adj1}[BC]_{N1}]D_{N2}]/[A_{Adj1}[B_{Adj2}[CD]_{N1}]]$) are taking separate incorporation movements, which is similar to cyclicity in classical lexicon. It is shown that the A-N compounding structures need to strictly follow the incorporation movement from Adj to Noun. All the other options, e.g. Adj occupies the specifier position of nP(N), are not tested successfully. This may be evidence to prove that A-N compounding structures still belong to classical “Word” instead of “Phrase”, conforming to relative

discussion about “Modifier-Noun” structure possessing many word characteristics in the literature.

The syntactic multiple-root incorporated structure is shown to be closely relevant with the framework of T3 tone sandhi in Mandarin Chinese as Concatenation rule, and stress assignment/asymmetric stress clash resolution in Shanghai Chinese.

Cyclic stress assignment rule in Shanghai Chinese is well represented in incorporated compounding structure, with correct prediction interacted with cyclic spell-out phases effects extended into the classical word domain. This potentially provides explanation to problematic imperative stress clash and stress clash avoidance in the literature.

T3 tone sandhi domain construction in Mandarin Chinese is potentially simplified into a simplified and unified framework of Concatenation rule after linearization of Morphological words, with the boundary of word domain no longer being problematic in classical literature, e.g., Optimality theory based MRU (minimal rhythm unit) framework in Chen (2009). This new framework also shows a noteworthy ability to deal with the exceptional situations in Chen (2009).

It is also shown that the concatenation algorithm of domain construction in T3 tone sandhi in Mandarin Chinese has a processing route that goes smoothly between the head positions of each phrase in the classical syntactic domain (e.g. V from VP, N from NP), **only** diving into the syntactic word structure in the case of compounding structures with equal or more than two morphemes, **not** in the monomorphemic structures. Under the current framework with phonological rules to be specific to T3 tone sandhi in Mandarin Chinese, the monomorphemic structures are opaque in the application process of some specific phonological rules. It is suggested to be a product or outcome of certain phonological rules’ processing economy and convenience. The specific algorithm of certain phonological rules shares a consistent concatenation process below and above the classical word domain, but does not apply in the large amount of monomorphemic structures.

This can be a potential reason to explain why we have a structurally-defined **word concept**: for some certain phonological rules, the opaque monomorphemic structures are structurally “left over”. Back to the dispute between lexicon-based word formation theory and syntactic word formation theory: **a.** The lexicon-based frameworks have been shown not to contribute enough to many phonological phenomena, e.g., phrasal lexical phonological rules. **b.** The

syntactic word formation theory seems still not to be capable of getting rid of "word" identity or structure completely yet. We may still get involved in the classical debate for some more time. Back to some discussion in the literature about traditional lexical phonological rules, phrasal phonological rules or some phonological rules vaguely floating between these two, the interaction of phonological rules below and above the classical word domain may be due to a structural problem—the processing results of some certain phonological rules form different structural domains/spaces, instead of a clear division between traditional lexicon and syntax.

References

Angeliki, Athanasopoulou. (2018). The Acquisition of Stress-Clash Resolution Strategies in Greek *Proceedings of the 42nd annual Boston University Conference on Language Development, ed. Anne B. Bertolini and Maxwell J. Kaplan, 45-57. Somerville, MA: Cascadilla Press.*

Ceccagno, Antonella and Basciano, Bianca. (2006). Classification, structure and headedness of Chinese compounds. *Lingue e Linguaggio V.2 (2006) 233–60.*

Chen, Mao-Hsu. (2018), Tone sandhi Phenomena In Taiwan Southern Min. *PUBLICLY ACCESSIBLE PENN DISSERTATIONS 3049. https://repository.upenn.edu/edissertations/3049.*

Chen, Matthew Y. (2009). Tone sandhi-Patterns across Chinese Dialects. *Cambridge University Press.*

Chomsky, N. (2001). Derivation by Phase. *In M. Kenstowicz (Ed.), Ken Hale: A Life in Language (pp. 1-52). Cambridge, MA: MIT Press.*

Duanmu, San. (1991). Stress and syntax–phonology mismatches: tonal domains in Danyang and Shanghai. *In D. Bates (ed.), Proceedings of the Tenth West Coast Conference on Formal Linguistics, Stanford Linguistic Association, pp.127–137.*

Duanmu, San. (1992a). End-based theory, cyclic stress, and tonal domains. *Proceedings of the Chicago Linguistic Society Meeting.*

Duanmu, San. (1993a). Rime length, stress, and association domains. *Journal of East Asian Linguistics 2, 1--44, 1993. Kluwer Academic Publishers. Printed in the Netherlands.*

Embick, David and Noyer, Rolf. (2007). Distributed Morphology and the Syntax/Morphology Interface. *G. Ramchand and C. Reiss eds., The Oxford Hand book of Linguistic Interfaces, Oxford University Press.*

Embick, David. (2010). Localism versus Globalism in Morphology and Phonology. *Linguistic Inquiry Monographs(60)*

Embick, David. (2013). Morphemes and morphophonological loci. *Distributed Morphology Today-Morphemes for Morris Halle. Editor(s):Ora Matushansky, Alec Marantz. URL: https://www.jstor.org/stable/j.ctt9qf7hr.10.*

Embick, David. (2013). Phase cycles, φ- cycles, and phonological (in)activity. *The Form of Structure, the Structure of Form, pp 271-286.*

Hayes, Bruce. (1990). Precompiled phrasal phonology. *The phonology-syntax connection, ed. by Sharon Inkelas and Draga Zec, pages 85–108.*

Harley, Heidi. (2010). Compounding in Distributed Morphology. *The Oxford Handbook of Compounding, Edited by Rochelle Lieber and Pavol Štekauer, Print Publication Date: Jul 2011 Subject: Linguistics, Language and Cognition Online Publication Date: Sep 2012 DOI: 10.1093/oxfordhb/9780199695720.013.0007.*

Halle, Morris. (1998). The Stress of English Words 1968–1998. *Linguistic Inquiry, Volume 29, Number 4, Fall 1998 539–568, by the Massachusetts Institute of Technology.*

Marantz, Alec. (1995). "Cat" as a phrasal idiom: Consequences of late insertion in Distributed Morphology.

Marantz, Alec. (2001). 'Words', *paper presented at the West Coast Conference on Formal Linguistics, University of Southern California Los Angeles, 24 February, 2001. Available online at http://web.mit.edu/marantz/Public/EALING/WordsWCCFL.pdf.*

Marvin, Tatjana. (2003). Topics in the Stress and Syntax of Words. *PhD dissertation in Massachusetts Institute of Technology.*

Marvin, Tatjana. (2013). "Is word structure relevant for stress assignment?" *in Alec Marantz and Ora Matushansky, eds. Distributed Morphology Today: Morphemes for Morris Halle, MIT Press, Cambridge, MA.*

Pak, Marjorie. (2008), The postsyntactic derivation and its phonological reflexes. *University of Pennsylvania PhD dissertation.*

Scott, Jackson and Jeffrey, Punske. (2013). Deriving English compound stress: insights from Distributed Morphology and multiple spell-out. *Linguistic Analysis 38 (3-4), 243-274*

Vigário, Marina. (2010). Prosodic structure between the prosodic word and the phonological phrase: Recursive nodes or an independent domain? *The Linguistic Review 27 (2010), 485–530, DOI 10.1515/tlir.2010.017.*