

The Productivity of Variable Disyllabic Tone Sandhi in Tianjin Chinese

Jie Zhang^a

Jiang Liu^b

^aDepartment of Linguistics, University of Kansas, USA

^bDepartment of Asian Languages and Literatures, University of Minnesota, USA

Running head: Tone Sandhi Productivity in Tianjin Chinese

Correspondence should be addressed to:

Jie Zhang

Department of Linguistics

The University of Kansas

1541 Lilac Lane

Blake Hall, Room 427

Lawrence, KS 66045-3129

Email: zhang@ku.edu

Phone: +1 785-864-2879

Fax: +1 785-864-5724

Appendix 1: Additional Stimuli Information

All statistical information for the stimuli was gleaned from a corpus of written Chinese compiled from online resources by Da (2004), which includes 258,852,642 total Chinese characters (each character is pronounced with one syllable in Chinese). The use of Da's corpus to approximate the frequencies of Tianjin words is primarily motivated by the lack of corpora on spoken Tianjin. But there are two additional justifications for this. First, Tianjin and SC are closely related Northern dialects of Chinese with considerable contact due to geographical proximity. There is hence a considerable overlap in vocabulary between them. Second, different Chinese dialects share the same orthographic tradition. Chinese speakers are thus diglossic in a sense. The Chinese that our Tianjin speakers read and write is essentially the same as written Standard Chinese. For these two reasons, the words selected from a written Chinese corpus reasonably approximate the knowledge of these words from our Tianjin speakers. Similar points have been made in Zhao and Jurafsky (2009) in defense of their use of an SC written corpus for a Cantonese study.

REAL disyllabic words were selected from the corpus using bigram information provided by Da based on two sub-corpora — one of news and one of general fiction. The two sub-corpora have a total of 28,278,285 bigrams. The mean raw bigram frequency for the high frequency words is 3,721, and that for the low frequency words is 178. One-way ANOVAs indicated that there is no significant difference in the log raw frequencies among different tonal combinations for either the high frequency ($F(5, 42) = 0.421, p = .832$) or low frequency words ($F(5, 42) = 1.498, p = .211$). Frequencies for the first syllables in high-frequency REAL, low-frequency REAL, and PSEUDO words include the frequencies of all homophonous characters, provided that

the characters are among the 3,500 most commonly used characters in Da's character corpus. In other words, these frequencies are approximations of the frequencies of the phonetic syllables with tones. For high-frequency REAL words, high-frequency first syllables have a mean raw frequency of 243,294, while low-frequency first syllables have a mean raw frequency of 70,284. For low-frequency REAL words, high-frequency first syllables have a mean raw frequency of 217,666, while low-frequency first syllables have a mean raw frequency of 58,067. For PSEUDO words, high-frequency first syllables have a mean raw frequency of 219,160, while low-frequency first syllables have a mean raw frequency of 57,666. One-way ANOVAs showed that for neither high-frequency nor low-frequency first syllables, is there a significant difference in the log raw σ_1 frequencies for different word types (high-frequency REAL, low frequency REAL, PSEUDO) (high frequency σ_1 : $F(2, 69) = 1.900$, $p = .157$); low frequency σ_1 : $F(2, 69) = 2.125$, $p = .127$). Within each word type, for neither high frequency nor low frequency first syllables, is there a significant difference in the log raw σ_1 frequencies for different tonal combinations (all $F(5, 18) < 1.7$, $p > .05$) except for one case: for high frequency REAL words, when the first syllable frequency is high, there is a significant effect for tonal combination ($F(5, 18) = 3.303$, $p = .027$), primarily caused by the higher σ_1 frequency of 4+1 than compared to 3+2.

To make the acoustic identification of the tone carrier easier, we primarily used obstruent onsets for the syllables in our stimuli, but occasional sonorant onsets were used for lexical frequency concerns. We did not actively control for aspiration in the obstruent onsets or vowel height in the rimes across different word types due to difficulties with frequency controls. But for every word type in every sandhi, there is typically a combination of aspirated and unaspirated obstruent onsets as well as different vowel heights. We thus did not expect these parameters to significantly influence on our f_0 results.

The full list of stimuli used in the experiment is given below. For high-frequency REAL (I), low-frequency REAL (II), and PSEUDO (III) words, within each base-tone combination, the four words above the thick solid line have a high-frequency first syllable, while the four words below the thick solid line have a low-frequency first syllable. The transcriptions represent the pronunciation of newer varieties of Tianjin, which have adopted the robust dental (ts, ts^h, s) vs. retroflex (tʂ, tʂ^h, ʂ) distinctions of Standard Chinese. Older varieties generally realized both series as dentals. Frequency calculations are based on the newer pronunciation.

I. High-frequency REAL words

| Base tones | Chinese bigram | IPA | Gloss | Bigram freq. | σ_1 freq. |
|------------|----------------|--------------------------------------|------------------|--------------|------------------|
| 1+1 | 今天 | tɕin t ^h jen | ‘today’ | 12,659 | 332,854 |
| | 飞机 | fei tci | ‘airplane’ | 3,475 | 280,231 |
| | 期间 | tɕ ^h i tɕjen | ‘period of time’ | 3,638 | 275,049 |
| | 资金 | tsi tɕin | ‘funding’ | 2,932 | 199,928 |
| | 增加 | tsəŋ tɕja | ‘to increase’ | 4,594 | 97,038 |
| | 专家 | tʂwan tɕja | ‘expert’ | 4,321 | 88,873 |
| | 双方 | ʂwan fən | ‘both ways’ | 4,132 | 72,559 |
| | 推出 | t ^h wei tʂ ^h u | ‘to launch’ | 1,171 | 82,512 |
| 3+2 | 保持 | pau tʂ ^h i | ‘to maintain’ | 3,286 | 227,852 |
| | 检察 | tɕjen tʂ ^h a | ‘to inspect’ | 2,187 | 158,877 |
| | 祖国 | tsu kwɔ | ‘motherland’ | 2,070 | 174,335 |
| | 组成 | tsu tʂ ^h əŋ | ‘to consist of’ | 1,963 | 174,064 |
| | 解决 | tɕje tɕje | ‘to solve’ | 6,937 | 41,762 |
| | 选择 | ɕjen tsɿ | ‘to choose’ | 2,797 | 82,498 |
| | 警察 | tɕiŋ tʂ ^h a | ‘police’ | 2,314 | 120,622 |
| | 考察 | k ^h au tʂ ^h a | ‘to investigate’ | 1,254 | 87,231 |
| 3+3 | 总统 | tsuŋ t ^h uŋ | ‘president’ | 8,159 | 184,470 |
| | 产品 | tʂ ^h an p ^h in | ‘product’ | 4,969 | 272,452 |
| | 减少 | tɕjen ʂau | ‘to reduce’ | 1,878 | 158,877 |
| | 保险 | pau ɕjen | ‘insurance’ | 1,658 | 227,852 |
| | 采取 | tʂ ^h ai tɕ ^h y | ‘to adopt’ | 3,918 | 27,717 |
| | 采访 | tʂ ^h ai fən | ‘to interview’ | 2,316 | 27,717 |
| | 选举 | ɕjen tɕy | ‘election’ | 2,054 | 82,498 |

| | | | | | |
|-----|----|------------|--------------------|--------|---------|
| | 整体 | tʂəŋ tʰi | ‘entirety’ | 1,185 | 103,560 |
| 3+4 | 表示 | pjau ʂi | ‘to express’ | 15,003 | 227,426 |
| | 保障 | pau tʂaŋ | ‘to safeguard’ | 2,745 | 227,426 |
| | 保证 | pau tʂəŋ | ‘to guarantee’ | 2,659 | 227,852 |
| | 体制 | tʰi tʂi | ‘system’ | 2,467 | 273,792 |
| | 恐怖 | kʰuŋ pu | ‘terrifying’ | 3,386 | 56,566 |
| | 比赛 | pi sai | ‘competition’ | 3,182 | 59,116 |
| | 解放 | tɕje faŋ | ‘to liberate’ | 2,905 | 41,762 |
| | 举办 | tɕy pan | ‘to hold’ | 1,983 | 75,727 |
| 4+1 | 必须 | pi cy | ‘must’ | 6,739 | 331,746 |
| | 信息 | cin ci | ‘information’ | 5,459 | 235,394 |
| | 战争 | tʂan tʂəŋ | ‘war’ | 4,751 | 408,786 |
| | 办公 | pan kuŋ | ‘to work’ | 4,617 | 226,789 |
| | 措施 | tsʰwɔ ʂi | ‘measure’ | 3,692 | 90,662 |
| | 斗争 | tou tʂəŋ | ‘to struggle’ | 2,433 | 84,087 |
| | 退休 | tʰwei cjou | ‘to retire’ | 918 | 52,528 |
| | 叹息 | tʰan ci | ‘to sigh’ | 774 | 61,933 |
| 4+4 | 电话 | tjen xwa | ‘telephone’ | 7,970 | 237,908 |
| | 报告 | pau kau | ‘to report’ | 5,457 | 268,590 |
| | 电视 | tjen ʂi | ‘television’ | 4,376 | 237,908 |
| | 爆炸 | pau tʂa | ‘to explode’ | 2,346 | 268,590 |
| | 促进 | tsʰu tcin | ‘to further’ | 3,463 | 30,970 |
| | 扩大 | kʰwɔ ta | ‘to expand’ | 2,684 | 81,713 |
| | 状态 | tʂwaŋ tʰai | ‘state of affairs’ | 1,714 | 95,916 |
| | 炸弹 | tʂa tan | ‘bomb’ | 1,024 | 41,260 |

II. Low-frequency REAL words

| Base tones | Chinese digram | IPA | Gloss | Bigram freq. | σ_1 freq. |
|------------|----------------|------------|----------------------|--------------|------------------|
| 1+1 | 疏忽 | ʂu xu | ‘to neglect’ | 103 | 262,768 |
| | 清官 | tɕʰiŋ kwan | ‘honest official’ | 54 | 329,953 |
| | 丰碑 | fəŋ pei | ‘monument’ | 39 | 230,581 |
| | 光纤 | kwaŋ cjen | ‘optical fiber’ | 27 | 147,639 |
| | 鞠躬 | tɕy kuŋ | ‘to bow’ | 189 | 64,352 |
| | 脱钩 | tʰwɔ kou | ‘to disconnect’ | 146 | 23,918 |
| | 宽松 | kʰwan suŋ | ‘loose fitting’ | 35 | 92,573 |
| | 刀疤 | tau pa | ‘knife scar’ | 8 | 27,292 |
| 3+2 | 死刑 | si ciŋ | ‘death sentence’ | 437 | 136,194 |
| | 铲除 | tʂʰan tʂʰu | ‘to eradicate’ | 162 | 272,452 |
| | 倒台 | tau tʰai | ‘to fall from power’ | 69 | 178,045 |
| | 简明 | tɕjen miŋ | ‘concise’ | 51 | 158,877 |
| | 躲藏 | twɔ tsʰaŋ | ‘to hide’ | 94 | 23,636 |

| | | | | | |
|-----|----|--------------|------------------------|-----|---------|
| | 涨幅 | tʃaŋ fu | ‘increase’ | 71 | 37,729 |
| | 匪徒 | fei tʰu | ‘bandit’ | 33 | 22,839 |
| | 掌勺 | tʃʰaŋ ʃau | ‘chef’ | 17 | 37,729 |
| 3+3 | 感慨 | kan kʰai | ‘to sigh with emotion’ | 483 | 267,307 |
| | 阻止 | tsu tʃi | ‘to prevent’ | 634 | 174,064 |
| | 辗转 | tʃan tʃwan | ‘to toss and turn’ | 152 | 164,916 |
| | 剪彩 | tɕjen tsʰai | ‘to cut the ribbon’ | 47 | 158,877 |
| | 躲闪 | twɔ ʃan | ‘to duck’ | 147 | 23,636 |
| | 谱写 | pʰu ɕje | ‘to compose’ | 87 | 77,029 |
| | 拷打 | kʰau ta | ‘to torture’ | 33 | 87,231 |
| | 搞垮 | kau kʰa | ‘to destroy’ | 24 | 36,277 |
| 3+4 | 阐述 | tʃʰan ʃu | ‘to expound’ | 336 | 272,452 |
| | 警戒 | ciŋ tɕje | ‘to guard’ | 290 | 120,622 |
| | 倒闭 | tau pi | ‘to go bankrupt’ | 123 | 178,045 |
| | 手铐 | ʃou kʰau | ‘handcuffs’ | 102 | 409,742 |
| | 广阔 | kwaŋ kʰwɔ | ‘broad’ | 528 | 87,900 |
| | 朴素 | pʰu su | ‘plain’ | 239 | 77,029 |
| | 喜鹊 | ci tɕʰɥe | ‘magpie’ | 88 | 86,958 |
| | 补钙 | pu kai | ‘to take calcium’ | 18 | 53,117 |
| 4+1 | 震惊 | tʃɔn tɕiŋ | ‘to amaze’ | 569 | 133,951 |
| | 振兴 | tʃɔn ciŋ | ‘to revitalize’ | 403 | 133,951 |
| | 弊端 | pi twan | ‘shortcoming’ | 92 | 331,746 |
| | 吊销 | tjau ɕjau | ‘to revoke’ | 74 | 158,370 |
| | 税收 | ʃwei ʃou | ‘tax’ | 750 | 79,035 |
| | 衬衫 | tʃʰən ʃan | ‘shirt’ | 249 | 103,492 |
| | 贵妃 | kwei fei | ‘imperial concubine’ | 112 | 59,601 |
| | 唱腔 | tʃʰaŋ tɕʰjaŋ | ‘vocal music’ | 37 | 32,851 |
| 4+4 | 塑造 | su tsau | ‘to shape’ | 252 | 227,217 |
| | 叙述 | ɕy ʃu | ‘to narrate’ | 342 | 156,678 |
| | 罢课 | pa kʰɿ | ‘students’ strike’ | 90 | 380,447 |
| | 特赦 | tʰɿ ʃɿ | ‘to pardon’ | 40 | 239,091 |
| | 配套 | pʰei tʰau | ‘to match in a set’ | 521 | 65,518 |
| | 帐户 | tʃaŋ xu | ‘account’ | 67 | 88,041 |
| | 挂帅 | kwa ʃwai | ‘to be in command’ | 51 | 23,827 |
| | 肾病 | ʃən piŋ | ‘kidney disease’ | 24 | 82,001 |

III. PSEUDO words

| Base tones | Chinese bigram | IPA | σ_1 freq. |
|------------|----------------|-----------|------------------|
| 1+1 | 东胚 | tʉŋ pʰei | 224,625 |
| | 风堆 | fəŋ twei | 230,581 |
| | 区今 | tɕʰy tɕin | 208,015 |
| | 虚踢 | ɕy tʰi | 229,779 |

| | | | |
|-----|----|---------------------------------------|---------|
| | 帮车 | paŋ tɕɿ | 64,218 |
| | 超趋 | tɕ ^h au tɕ ^h y | 56,700 |
| | 居歌 | tɕy kɿ | 64,352 |
| | 翻屋 | fan wu | 52,604 |
| 3+2 | 保门 | pau mən | 227,852 |
| | 场煤 | tɕ ^h aŋ mei | 210,348 |
| | 管叠 | kwan tje | 201,310 |
| | 海菩 | xai p ^h u | 209,302 |
| | 比回 | pi xwei | 59,136 |
| | 纺池 | faŋ tɕ ^h i | 53,179 |
| | 苦凉 | k ^h u ljaŋ | 63,194 |
| | 省肥 | ɕəŋ fei | 58,642 |
| 3+3 | 体母 | t ^h i mu | 273,792 |
| | 显铁 | cjɛn t ^h je | 179,592 |
| | 腐满 | fu man | 144,463 |
| | 等躲 | təŋ twɔ | 266,296 |
| | 否纸 | fou tɕi | 64,772 |
| | 轨挺 | kwei t ^h iŋ | 40,512 |
| | 姐巧 | tɕje tɕ ^h jau | 41,762 |
| | 卡酒 | k ^h a tɕjou | 52,244 |
| 3+4 | 表副 | pjau fu | 226,768 |
| | 总特 | tsuŋ t ^h ɿ | 184,470 |
| | 百配 | pai p ^h ei | 139,693 |
| | 产困 | tɕ ^h an k ^h un | 272,452 |
| | 补妾 | pu tɕ ^h je | 55,244 |
| | 恐练 | k ^h uŋ ljen | 56,566 |
| | 审炸 | ɕən tɕa | 61,455 |
| | 左库 | tswɔ k ^h u | 51,946 |
| 4+1 | 触刊 | tɕ ^h u k ^h an | 224,620 |
| | 电葱 | tjen ts ^h uŋ | 237,908 |
| | 共掐 | kuŋ tɕ ^h ja | 215,777 |
| | 户呆 | xu tai | 208,731 |
| | 拜新 | pai cin | 60,605 |
| | 创吃 | tɕ ^h waŋ tɕ ^h i | 63,156 |
| | 顿标 | tun pjau | 63,693 |
| | 霎专 | ɕa tɕwan | 53,169 |
| 4+4 | 瘦棒 | ɕou paŋ | 272,962 |
| | 控静 | k ^h uŋ tciŋ | 204,203 |
| | 诉万 | su wan | 227,217 |
| | 特炮 | t ^h ɿ p ^h au | 239,091 |
| | 跪恰 | kwei tɕ ^h ja | 59,601 |
| | 抗界 | k ^h aŋ tɕje | 50,658 |

| | | | |
|--|----|-----------------------|--------|
| | 盼令 | p ^h an liŋ | 69,473 |
| | 善秀 | ʃan ɕjou | 67,091 |

IV. NOVEL words

| Base tones | Chinese bigram | IPA |
|------------|----------------|---------------------------------------|
| 1+1 | fo 多 | fwo twɔ |
| | neng 归 | nəŋ kwei |
| | nong 揪 | nuŋ tɕjou |
| | huai 吩 | xwai fən |
| 3+2 | pan 胡 | pan xu |
| | xia 林 | ɕja lin |
| | kuang 鱼 | k ^h waŋ ɥy |
| | kui 伦 | k ^h wei lwən |
| 3+3 | pei 犬 | p ^h ei tɕ ^h ɥən |
| | sei 法 | sei fa |
| | chua 轨 | tɕ ^h wa kwei |
| | pen 海 | p ^h ən xai |
| 3+4 | nüe 半 | nɥe ban |
| | tuan 路 | t ^h wan lu |
| | kuo 费 | k ^h wɔ fei |
| | pang 彻 | p ^h ɑŋ tɕ ^h ɣ |
| 4+1 | teng 颠 | t ^h əŋ tɕjən |
| | min 枯 | min k ^h u |
| | nang 掐 | nɑŋ tɕ ^h ja |
| | niu 班 | njou pan |
| 4+4 | diu 叫 | tjou tɕjau |
| | mang 赠 | mɑŋ tsəŋ |
| | ka 告 | k ^h a kau |
| | ting 榨 | t ^h iŋ tɕa |