Stuck in the middle: Mandarin medials in articulation, parsing, and association

James Myers
National Chung Cheng University
Lngmyers@ccu.edu.tw
2014/5/9

Abstract:

Is the Mandarin medial (prenuclear glide) represented mentally as part of the onset or the rime? I argue that three forces influence the apparent affiliation of the medial across different sources of evidence: articulation encourages the grouping of medials with onsets, parsing encourages the grouping of medials with rimes, and association strength modulates the effects of the other two forces. After showing that this framework explains the otherwise puzzling diversity of results in the previous literature, I describe an acceptability judgment experiment in which all three forces were minimized. The results revealed a consistently greater interaction between medials and onsets than between medials and rimes. Assuming that all three forces were truly minimized in the experiment, and that there are no other forces to consider, these results provide the most reliable evidence yet that the Mandarin medial may be represented mentally as part of the onset.

Keywords:

Mandarin, syllable structure, phonetics, acceptability judgments, experiment

1. Introduction

Chinese phonologists have traditionally analyzed the syllable into the onset (initial) and final, with the latter consisting not just of the rime, but of an optional prenuclear medial as well. This three-part onset-medial-rime division (though not the medial-rime grouping) underlies the Zhuyin Fuhao phonetic orthography, developed in China in the early twentieth century and still used in teaching and computer key-in systems in Taiwan. Thus the word [tɕʰiaŋ⁵⁵] "gun", *qiāng* in Pinyin romanization, is transcribed in the Zhuyin system with the three symbols ㄑㄧㄤ, where the first represents the onset /tɕʰ/, the third the rime /aŋ/, and the second the medial /i/ (consistent with the Zhuyin system, which I use in the experiment described below, I transcribe the Mandarin medials /i/, /u/ and /y/ as vowels, though nothing theoretical hinges on this choice).

The traditional grouping of the medial with the rime has retained its advocates into the generative era (e.g. Cheng 1973), but the evidence for it is equivocal. While it seems to be supported by certain psycholinguistic experiments (Wang and Chang 2001, Chang 2012), other arguments seem to show that the medial groups instead with the onset (Duanmu 2007), and the evidence from speech errors is mixed (Wan 1997). One response to this diversity of results is to conclude, with Yip (2003), that the medial is intrinsically grouped neither with the onset nor the rime. Yet even if this null hypothesis can be maintained, it wouldn't explain the conflicting results. A more productive approach, I think, is to follow the lead of Berg and Koops (2010), who suggest that conflicting evidence regarding syllable structure actually implies the existence of competing forces.

In the case of the Mandarin medial, I argue that the competing forces are articulation, which favors onset-medial grouping, and parsing, which favors medial-rime grouping. Both of these forces are modulated by a third, the association strength between syllable elements. After showing that the onset-medial and medial-rime association strengths are virtually identical in Mandarin, I describe an experiment designed to minimize the articulatory and parsing forces as well. The results revealed a greater interaction between onsets and medials than between medials and rimes. Assuming that all three forces were truly minimized in the experiment, and that there are no other forces to consider, these results would seem to provide more reliable evidence than ever before that the Mandarin medial may be represented as part of the onset.

I begin in section 2 by reviewing the *a priori* arguments for and against the null hypothesis that the Mandarin medial is intrinsically unaffiliated with both onset and rime, supplementing these arguments with new evidence from association strength. Then in section 3 I reanalyze the arguments that have been made in the literature for grouping the medial with the onset or rime, showing how articulation, parsing, and association strength play key roles. In section 4 I describe an experiment using a speeded acceptability judgment task of written nonlexical syllables, which does not require articulation or parsing. Section 5 provides a conclusion.

2. Dependency and association strength

One argument for affiliating the Mandarin medial within one of the adjacent units is that universal syllable theory posits an onset and a rime, but nothing in between. This argument against an unaffiliated medial has been presented most forcefully by Duanmu (2007, 2010), who concludes that at a suitable level of abstraction, all languages have exactly the same syllable structure. Details aside, the core of the argument is simplicity: if languages like English don't have a separate medial unit, then the overall theory would be simpler if Mandarin didn't either.

However, even if the medial is indeed universally situated within the onset or the rime, its

apparent affiliation may be affected not only by its hierarchical constituency, but also by its interactions with other elements. The latter mechanism has a long history, formalized most notably as dependency (Anderson and Ewen 1987). The appeal of the dependency notion here is that it offers a way in which the Mandarin medial could seem to switch its affiliation between onset and rime, due to whichever force is most active in a particular source of evidence.

It turns out to be very difficult to distinguish hierarchical constituency from dependency. To do so, one needs to detect whether speakers have a representational unit independent of the constituent's elements. Thus to test if the English rime is a hierarchical constituent, Kapatsinski (2009) trained English speakers to associate consonant-vowel sequences (bodies) and vowel-consonant sequences (rimes) in CVC syllables with "affixes" (treating all of the materials as if they came from some imaginary language). Crucially, the body and rime associations were designed to conflict with those for the individual consonants and vowels themselves. For example, one group of participants was trained to associate /æ/ and /ʃ/, separately, with the suffix /num/, but at the same time to associate the rime /æʃ/ with the suffix /mɪn/, while another group was trained on /ʃ/-/num/ and /æ/-/num/ segmental associations but the /ʃæ/-/mɪn/ body association. After confirming that the segment-affix associations were learned, Kapatsinski found that only the rime-affix associations could be learned on top of them, not the body-affix associations. This suggests that English speakers do indeed have an independent rime unit.

While Kapatsinski's results seem to establish the status of English rimes as true hierarchical constituents, his method is unfortunately too complex to apply to the Mandarin medial issue. His full experiment involved far more than the two conditions sketched above, and the highly restrictive nature of Mandarin phonotactics makes it impossible to include all of the controls needed to rule out alternative learning strategies in his paradigm. Hence the best we can do is to try to understand the forces that influence dependency to see how they might be experimentally minimized.

A standard way to quantify the lexical regularities that give rise to dependencies in a learner's mind is to compute the statistical association strength of elements. This can be done in various ways, but Lee and Goldrick (2008) argue that the best measure is something called r_φ (r-phi), which represents the (ordinary Pearson) correlation between two binary variables. In this case, the binary variables represent the presence/absence of elements in a syllable. For example, to measure the association strength between /t/ and /a/ in English syllables, we list all English syllables, and if a /t/ is present, we mark that syllable with 1, otherwise with 0, and similarly for /a/, and then compute r_φ for these two sets of 0s and 1s.

Lee and Goldrick (2008) apply this measure to probe the structure of syllables in English, which has an onset-rime structure, and Korean, which, highly unusually, seems to have a body-coda structure (see e.g. Yoon and Derwing 2001, Berg and Koops 2010). Since they are interested in the magnitude of association strength and not its sign, they use the absolute value of r_φ: larger values (up to a maximum of 1) indicate that the lexicon "pays attention" to a particular combination, mostly encouraging it or mostly restricting it. They found that the association strength in English syllables is significantly greater for rimes than for bodies, while the reverse is the case for Korean syllables.

Berg and Koops (2010) provide further evidence for the body-coda structure of Korean syllables, but fail to find a difference in rime versus body association strengths. Their struggle to explain this anomaly leads them to propose that association strength competes with a synchronic force favoring rimes (which I will identify below with parsing). However, while association strength may be informative to learners, it is synchronically a fixed property of the lexicon. Moreover, without an association strength asymmetry, Berg and Koops provide no plausible way Korean speakers could learn their typologically unusual syllable structure,

merely pointing to the left-branching structure of Korean syntax (though most languages with left-branching syntax, like Japanese, have right-branching syllables). More likely, their association strength results differ from those of Lee and Goldrick (2008) because they used different statistical methods; absence of evidence for an association strength asymmetry in Korean syllables doesn't necessarily imply evidence of absence.

What about the Mandarin medial? Using the lexical Mandarin syllables listed in Tsai (2000), I calculated $|r_\varphi|$ (the absolute value of r_φ) for all combinations of the syllable elements assumed by Zhuyin Fuhao (twenty-one onsets, twelve rimes not counting the rhotic rime, which is never combined with other elements tautomorphemically, and the three medials /i/, /u/, and /y/). For example, the onset-medial combination with the lowest association strength is /ny/ (.003) and that with the highest is /xy/ (.253), in part because velar onsets can never occur with front vowels, while for medial-rime combinations these extremes are /in/ (0.003) and /iɛ/ (.244), in part because the rime /ɛ/ can only follow front vowels (note, again, that association strength reflects both negative and positive interactions). The overall results are summarized in Table 1.

Table 1. Association strengths ($|r_\varphi|$) for Mandarin syllable elements

Combination	Mean	Median	Minimum	Maximum
Onset-Medial	.086	.061	.003	.253
Medial-Rime	.083	.063	.003	.244
Onset-Rime	.021	.016	< .001	.125

The association strengths are significantly different across the three combinations by the Kruskal-Wallis test, which does not depend on distribution shape ($\chi^2(2) = 22.95$, $p < .001$), but this is due solely to the onset-rime grouping being different from the other two, as shown by the similarly distribution-free Mann-Whitney test, used by Lee and Goldrick (2008) in their analyses (onset-medial vs. medial-rime: $U = 1,317$, $p = .18$; onset-medial vs. onset-rime: $U = 14,883$, $p < .001$; medial-rime vs. onset-rime: $U = 7,786$, $p < .001$).

Thus unlike the English or Korean nucleus, the Mandarin medial has identical association strengths with both adjacent elements. Lexical statistics do not inform a Mandarin learner about the medial's affiliation.

3. Explaining the medial muddle

If association strength is agnostic on the position of the Mandarin medial, why does the literature suggest that it is grouped with the onset (according to some evidence) or with the rime (according to other evidence)? My answer builds on the banal notion that the implications of a piece of evidence depend on the intrinsic nature of that evidence. We generally do not know ahead of time which source of evidence (if any) provides the least distorted image of the linguistic representations we posit, making it necessary to probe in multiple ways (Derwing et al. 2011 apply this approach to the syllable structure of another Sinitic language, Southern Min). At some point, however, it becomes crucial to figure out *why* different evidence sources yield different results. At least in the case of conflicting evidence about the Mandarin medial, I suspect that the key explanatory variables are articulation and parsing.

3.1 Articulation

When we look at the arguments that have been given for grouping the medial with the

onset, most notably by Duanmu (2007, 25-30, 79-81), we see a curious preponderance of articulatory evidence. Duanmu's first argument is that the medial is coarticulated with the onset, and his second argument is overtly articulatory as well: the onset-medial combinations attested in Mandarin are always expressible in terms of secondary articulations (Ci as C^j with [-back], Cu as C^w with [+round], and Cy as $C^ɥ$ with [-back, +round]), and gaps in possible combinations are predictable from articulatory constraints. Thus retroflexes cannot cooccur with medials because the latter's [+high] feature conflicts with retroflection, and labials cannot cooccur with round medials or velars with front medials because these combinations require a single articulator to be associated with two features ([-continuant] plus [+round] or [-back]). In Duanmu's analysis, higher-level cognitive constraints like the Obligatory Contour Principle (OCP) are not relevant; Mandarin does not ban $/b^w/$ because both /b/ and rounding invoke the Labial node (cf. McCarthy 1988), but rather because the articulation of $/b^w/$ would require simultaneous stopping and rounding. Of course this is not purely a matter of physics (/bua/ is a legal syllable in the fellow Sinitic language of Southern Min), but Duanmu implies that the Mandarin speaker's knowledge of this constraint is encoded in terms of articulatory parameters.

Duanmu's third and fourth arguments are articulatory in the same sense: the third notes that the medial has no effect on syllable duration, as expected if the onset and medial are articulated as a single segment, and the fourth notes that varieties of Mandarin vary in the degree of onset-medial coarticulation, so for example, the morpheme for 'bird' can be pronounced anywhere from [niau] to [ɲau]. Again, these are language-specific patterns, but encoded in terms of articulation.

His final argument notes that the medial cannot cooccur with what he calls the zero onset (a variously realized filler that creates an onset in lexically onsetless syllables), so they must therefore compete for the same slot. However, the zero onset is of course unnecessary if a medial is present to give the initiation of syllable articulation a suitably tight degree of stricture on its own. The tendency for syllables to favor initial oral stricture is often described in terms of the acoustic-perceptual property of sonority, but sonority actually has a close connection with the degree of opening in the vocal tract (see e.g. Beckman et al. 1992).

Duanmu acknowledges that evidence from other sources, such as language games, do not point clearly towards grouping the medial with the onset, but dismisses such evidence as ambiguous. It is nevertheless intriguing that all of his own positive arguments seem to have an articulatory interpretation. This suggests that even if Duanmu's onset-medial grouping is a genuine hierarchical constituent of the Mandarin syllable, it may be motivated in part by articulatory forces.

Why might articulation favor such a grouping? One clue comes from articulatory dynamics (e.g. Browman and Goldstein 1988). Roughly, the onset of the syllable initializes rhythms associated with each of the articulators, with those for the initial consonant and vowel starting out in phase, whereas articulatory coordination later in the syllable depends in part on the rhythms set up at this initiation point. Such dynamics would, at the very least, make it easier for languages like Mandarin to phonologize the coordination of medials with onsets than with rimes.

One might expect articulatory forces to be dominant in speech errors as well. However, speech errors actually involve cognitive processes (manipulation of phonological elements) that occur prior to the initiation of articulation (Levelt 1989). Speech errors nevertheless provide insights into syllable structure since simple errors (involving only one erroneous operation) are expected to be more common than complex ones (involving more than one operation or multiple elements). Thus if two subsyllabic units tend to participate together in speech errors, we can infer that the processing system treats them together.

Surprisingly, Wan (1997) reports that the treatment of the medial depends on the place of

articulation of the onset. For onsets that are labial, dental, or retroflex, which she calls anterior, the medial groups with the rime, while for onsets that are palatal or velar, which she calls posterior, the medial groups with the onset. This curious pattern may be explainable in terms of association strengths. As summarized in Table 2, there is no overlap across Wan's two onset classes: the anterior onset with the maximum mean medial association is the retroflex /tʂʰ/ (.089), and the posterior onset with the minimum mean medial association is /k/ (.091).

Table 2. Mean association strengths ($|r_\varphi|$) for all medials with the onset categories of Wan (1997)

Onset	Mean	Median	Minimum	Maximum
Anterior	.063	.055	.036	.089
Posterior	.142	.166	.091	.190

The reason medial association strengths for palatal onsets are closest to those for velars (forming Wan's posterior group) may be because these two places are in complementary distribution with respect to the medials (though see Cheng 1973 and Duanmu 2007 for complexities): velars cooccur with /u/ but not /i/ or /y/, while palatals cooccur with /i/ and /y/ but not /u/. In the traditional three-part syllable structure, in fact, palatals are actually obliged to cooccur with /i/ or /y/.

In short, in speech production, the medial seems to be articulated with the onset, but in speech errors, which are generated prior to articulation, the onset-medial grouping only appears if there is a high association strength between these elements.

3.2 Parsing

While the Mandarin medial tends to be pulled towards the onset by articulation, it is also pulled towards the rime by another force: parsing. Medial-rime groupings create structures in which the last-added element (when processing the structure from beginning to end) is more strongly associated with the element added just previously.

In the parsing of the hierarchical structures of syntax, there is just such a principle favoring right-branching structures. Among other things, this principle explains why in the sentence *Joe said that Martha expected that it would rain yesterday*, the adverb *yesterday* tends to be interpreted as associated with *it would rain*, not with *Joe said* or *Martha expected* (Kimball 1973, 27). This principle is apparently motivated by working memory constraints, since it is easiest to associate a new piece of information with a recent piece of information still active in memory.

Even though phonology is not like syntax in every way, working memory constraints should be relevant to both, so we expect right-dependent structures to be favored when parsing syllables as well. Berg and Coops (2010) come to a similar conclusion, independent of the above syntactic argument, by considering the cohort model of spoken word recognition (Marslen-Wilson and Tyler 1980). According to this model, as a spoken word unfolds in time, the listener's mental cohort of lexical candidates consistent with what has been heard so far shrinks as more information comes in, eventually leaving only the intended word.

Thus by the time the listener reaches the next-to-last phonological element, the final element is already highly predictable. This suggests two things. First, as Berg and Coops (2010) argue, knowing the association strength of the current element with other elements in the inventory helps the listener guess the next element, but this information is more helpful towards the end of the syllable, when competing candidates are fewer. Second (going beyond

Berg and Coops), repeated experiences of correctly predicting the final element from the next-to-last one may increase the dependency between these two final units. This asymmetry in dependency may be even greater if the inventory of syllable-final elements is smaller than syllable-initial elements (as is the case in Mandarin).

If the preference for right-dependent structures is due to working memory constraints during parsing, we predict that the Mandarin medial will tend to associate with the rime in processes that are particularly dependent on working memory. This may include the speech error evidence of Wan (1997), since such errors are made in working memory, prior to articulation. Indeed, in most of her cases (i.e. when onset-medial association strength is low), the medial seems to group with the rime.

Working memory also seems to play an essential role in the language-game-like tasks that have been used experimentally to study the position of the Mandarin medial. I agree with Yip (2003) and Duanmu (2007) that conventionalized language games do not provide useful evidence, in this case because if both onset-medial and medial-rime dependencies can be processed at all, games can be invented that use either. In the experimental language games, by contrast, the goal is to discover what speakers spontaneously treat as the more natural grouping.

Wang and Chang (2001) and Chang (2012) both report this type of experiment, where participants are asked to merge two Mandarin syllables into one or to split one syllable into two. Practice items contain no medials, but the key test items do; the question is whether the medials will be deleted or preserved along with the adjacent onset or with the adjacent rime. Thus in order to perform such tasks, participants must hold one or two syllables in memory while they construct the merged or split forms, encouraging their parsing.

In fact, both studies did find a strong preference for medial-rime groupings. This preference was far from complete, however: between a fourth and a third of the responses indicated onset-medial groupings instead. Moreover, since Chang (2012) designed his materials and analyses to test for the influence of association strength, he was able to show that items with greater onset-medial association strength showed an increased tendency to yield onset-medial responses. Thus language-game experiments suggest that the force pushing responses towards medial-rime groupings (here suggested to be parsing) competes with other forces, like association strength, favoring onset-medial groupings. Even the possibility of a hierarchical onset-medial constituent remains open, given the surprisingly large proportion of onset-medial responses.

Of course, working memory also plays a role in experimental tasks besides syllable manipulation. This may explain why support for a medial-rime grouping was also reported by Li (2009) in an acceptability judgment experiment, where participants had to decide how Mandarin-like nonlexical syllables were. In order to extract information about syllable structure, Li analyzed his judgment data using computational models that grouped the medial with the onset, with the rime, or with both, and tested which of these three models fit his data best.

More precisely, he first entered lexical Mandarin syllables, as strings of phonological feature bundles, into the acquisition model of Hayes and Wilson (2008), which induces phonotactic constraints and their strengths by counting feature combinations in the training data (via maximum entropy). The Hayes and Wilson model also allows the researcher to specify "innate" constituent structures, which determine which feature combinations are possible targets for phonotactic constraint induction. In Li's onset-medial model, a boundary was posited between the medial and the rime, so that no medial-rime phonotactic constraints could be learned, and similarly for the medial-rime model. Only the medial-rime model generated acceptability scores that were significantly (and positively) correlated with those of the actual speakers.

Unlike language games, acceptability judgments do not necessarily burden working memory or require much parsing. However, the specific methods used in Li (2009) may have encouraged parsing anyway. First, judgments were made on a nine-point scale (1 = most like Mandarin, 9 = least like Mandarin), and such a wide range of choices may have encouraged participants to recruit working memory to store test syllables during decision-making and to compare the current response with earlier ones for consistency across trials. Second, following Bailey and Hahn (2001), Li included lexical syllables among his test items, which may have further encouraged memory retrieval processes, since if an item is lexical, the best response is to give it a full score (i.e. 1). Third, each syllable was presented not just visually (in Zhuyin Fuhao), but also auditorily; spoken syllables unfold in time, again encouraging parsing.

This survey suggests that if we want evidence for the default affiliation of the Mandarin medial, we must be careful to use an experiment that invokes neither articulation nor working-memory-driven parsing.

4. An acceptability judgment experiment

Although we have seen that an acceptability judgment task is not guaranteed to eliminate the influence of parsing, I reasoned that it should still be possible to minimize this influence under the right conditions. Thus I ran a new analysis of experimentally collected Mandarin syllable acceptability judgments. The experiment was not designed to address the medial question specifically, but is instead what is known as a megastudy (Balota et al. 2012), where a very large and varied sample of experimental data is collected and then analyzed freely, like a corpus. The present experiment collected native speaker acceptability judgments on all logically possible nonlexical syllables in Mandarin.

Unlike the syllable acceptability judgment experiment of Li (2009), participants only had to make a binary choice: like or unlike Mandarin. No overt comparisons with lexical syllables were necessary either, since only nonlexical syllables were tested, and the much greater number of test items (ten times more than tested by Li) presumably made it even harder for participants to maintain memory traces of individual test items. Finally, the stimuli were presented only visually (in Zhuyin Fuhao), making it less useful for participants to parse. Nevertheless, we expect judgments of these written stimuli to depend primarily on phonology, not orthography. This has already been demonstrated for nonword judgments in written English (Bailey and Hahn 2001), and the Zhuyin system is even more transparent. Moreover, unlike written English, Zhuyin Fuhao is used exclusively for encoding pronunciation (in teaching, dictionaries, and computer key-in systems).

In order to study how the participants naturally grouped syllable elements together, I computed syllable element frequencies for each test item, one each for the item's onset, medial, and rime, representing the number of lexical syllables (ignoring tone) that share that element with the test item. For example, for the test item ㄅㄧㄚˋ /pia^{51}/, its onset frequency is the number of lexical Mandarin syllables with the onset /p/ (56 in Tsai 2000), its medial frequency is the number with the medial /i/ (278), and its rime frequency is the number with /a/ (94). For test syllables missing one or more of these elements, that element frequency for the test item was set to zero (I come back to this point later). While element frequency encodes a primitive kind of phonotactic probability (i.e. the probability of that element appearing in the lexicon), finding an interaction of the frequencies of the elements X and Y in their effect on judgments would imply that X and Y must be processed together, in some sense.

Thus the goal of this analysis was to find out whether acceptability judgments are affected by onset-medial interactions, by medial-rime interactions, by neither, or by both.

4.1 Methods

4.1.1 Participants

114 native speakers of Mandarin Chinese, undergraduates at a university in Taiwan, were paid for their participation. The results for four participants were excluded because they failed to return to the second session, leaving 110 participants for the analysis.

4.1.2 Materials

Each stimulus consisted of a sequence of Zhuyin Fuhao symbols in the prosodically permitted order of (optional) onset consonant, (optional) medial, and rime, along with a tone, written horizontally, as in computer key-in systems. This gave 4,516 logically possible syllables. We then excluded the 1,239 lexical syllables (with full tones and ignoring the highly restricted rhotic rime) listed in Tsai (2000), along with three recent neologisms (*biàng*, *liāng*, *biāng* in Pinyin). The total number of nonlexical test syllables was thus 3,274.

Element frequencies were calculated separately for onsets, medials, and rimes, based on the syllable types in Tsai (2000); their distributions were reasonably normal and did not need to be logarithm-transformed (cf. Baayen 2008). Following a reviewer's suggestion, the analyses also took into account the onset-medial and medial-rime association strengths ($|r_{\varphi}|$). Even though these two measures are not significantly different in Mandarin, speakers might pay more attention to one than the other.

However, the reliability of the analysis requires that the potential predictor variables (here, the onset, medial, and rime frequencies and the onset-medial and medial-rime association strengths) not be collinear (confounded) across the observations that we will analyze (here, all valid responses to all test items by all participants). The overall collinearity measure κ was a respectably low 12.4 ($\kappa > 30$ would indicate harmful collinearity; Baayen 2008), but there were still correlations between the two association strengths ($r^2 = .10$, $p < .001$, implying that 10% of the variance in one is predictable from the other) and between medial frequency and each association strength (onset-medial: $r^2 = .24$, $p < .001$; medial-rime: $r^2 = .21$, $p < .001$). These correlations are not theoretically important (the association strengths and medial frequency all involve medial counts), but they may potentially muddy the statistics. Moreover, the element frequencies had wide ranges while the association strengths had narrow ranges, making the statistical algorithm more likely to crash.

The solution to the first of these problems was to use a method called orthogonalization, replacing each association strength with the residuals derived from a (linear regression) model predicting each from the three element frequencies, yielding variables which represent all of the variance in the association strengths not predictable from the element frequencies. This entirely eliminated the correlations between association strength and medial frequency, while greatly reducing the correlation between the two association strengths ($r^2 = .01$, $p < .001$), without losing much in the bargain (each transformed measure still accounted for 75% of its original variance). The solution to the second problem was to standardize all five predictors (using z scores) to put them all on the same scale with affecting the interpretation of interactions (Aiken and West 1991). Standardization also eliminated all remaining traces of collinearity ($\kappa = 1.1$).

4.1.3 Procedure

The experiment was run using E-Prime 2.0 (Schneider et al. 2002). After conducting some pretests (not relevant here), all participants were visually presented with all syllables in

random order in two blocks (1,599 in the first, 1,675 in the second; there were rest breaks every 160 trials). Participants were asked to judge each stimulus as "like Mandarin" (*xiàng Guóyǔ*) by pressing the "L" key (right hand) or as "not like Mandarin" (*bú xiàng Guóyǔ*) by pressing the "S" key (left hand). Each trial began with the mark '+' for 1000 milliseconds (one second) to ensure that their eyes were pointed in the right direction when the stimulus appeared, followed by the target syllable for 4000 ms. A trial ended when the participant made a response or the 4000 ms limit was reached.

The task was split into two blocks because of its sheer length: it took participants three hours to complete the pretests and judge all syllables. The second session took place on the same day (after a two-hour lunch break), or on another day (one to ten days later).

4.2 Results and discussion

1,727 responses were not recorded for being too slow, and an additional 12,209 were dropped for being implausibly fast (less than 100 ms). Since there were 360,140 (= 3,274 × 110) possible responses, the missing data comprised less than 3% of the total. Response latencies were typical for reading experiments, though with greater variability (mean 802 ms, standard deviation 587 ms). By far the most common response was to reject items as not like Mandarin; only 11% of the valid responses were positive, and one participant even rejected all items (as noted by Myers & Tsay, 2005, the small syllabary of Mandarin sharpens the perceived boundary between lexical and non-lexical items). Nevertheless, given the large number of items and participants, this still means that there were 37,603 positive responses from the 109 remaining participants.

Since the dependent variable involved binary responses grouped by participant and item, I modeled the results with mixed-effects logistic regression (see e.g. Baayen 2008). Barr et al. (2013) recommend that treatment-by-participant interactions should also be modeled, not just overall participant differences, but as they also acknowledge, doing so tends to cause the mixed-effects logistic regression algorithm to crash. Thus I had no choice but to neglect their recommendation.

The regression analysis modeled the probability of positive acceptability responses as a function of the onset, medial, and rime frequencies and all of their possible interactions, along with onset-medial and medial-rime association strengths (no interactions with the last two were tested, since this would merely complicate the model to no theoretical end). The key results are illustrated in Figure 1, which shows how acceptability is affected by onset and rime frequency, each as modulated by the three medial frequencies. The acceptability scores are all negative, since more syllables were judged unacceptable than acceptable, but higher values (higher in the plots) still represent higher acceptability.

Note first that overall acceptability for the medial /u/ is a little higher than that for the medial /i/ and much higher than that for the medial /y/. These ratings almost perfectly match the relative frequencies of these medials: /u/ is found in 340 lexical Mandarin syllables, /i/ in 278, and /y/ in only 74. There are also overall positive effects of onset frequency (as shown by the slope of the lines in the onset plot on the left) and of rime frequency (in the plot on the right).

Crucially, the figure also suggests that onset frequency has a stronger effect on judgments in the presence of a higher-frequency medial. Thus the two upper lines (for /u/ and /i/) in the onset plot rise more steeply than the bottommost line (for /y/), particularly for /i/. If there were no onset-medial interactions, these three lines would be parallel. The plot for rime frequencies does not seem to show such an interaction.
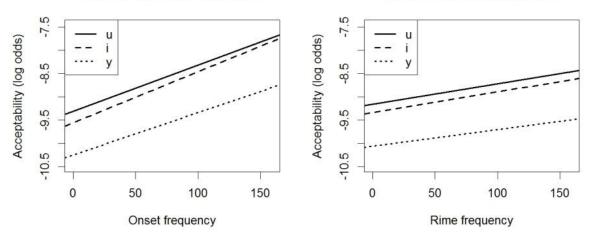
Figure 1. Acceptability as a function of onset, medial, and rime frequencies. The lines represent by-item linear fits (equivalent to logistic regression fits due to the log odds transformation in the y-axis).

These impressions, along with other relevant findings, are statistically confirmed by the mixed-effects logistic regression model, summarized in Table 3. The standardized coefficients (β) indicate the effect sizes for various predictors and their interactions, with the signs corresponding to slope direction. The large negative coefficient for the intercept merely confirms that the vast majority of responses are negative, regardless of the item type. More relevantly, all three element frequencies show significant positive effects, consistent with previous research on the positive influence of phonotactic probability on acceptability (e.g. Bailey and Hahn 2001 for English, Myers and Tsay 2005 for Mandarin). Both onset-medial and medial-rime association strengths have significant effects as well, though in a negative direction, in part because they include element combinations that are disfavored in the Mandarin lexicon. Their virtually identical coefficients are not significantly different by a likelihood ratio test comparing the original model with one assuming their identity ($\chi^2(1)$ = 0.014, $p > .9$). Thus Mandarin speakers seem to pay equal attention to onset-medial and medial-rime association strengths, consistent with the equality of these measures in the Mandarin lexicon.

The model also confirms an asymmetry in the element frequency interactions. While there is a statistically significant positive interaction for onset frequency × medial frequency, consistent with the enhancement effect seen in the onset frequency plot, there is no significant interaction for medial × rime (nor for onset × rime and onset × medial × rime; the marginal p for the former is not reliable given my inability to follow the recommendation of Barr et al. 2013). In other words, when Mandarin speakers (unconsciously) consider the influence of onset frequencies on acceptability, they take medial frequencies into account and vice versa, but rime frequencies do not modulate the influence of medial frequencies or vice versa. If this statistical analysis can be trusted, we have uncovered a tendency for Mandarin speakers to group medials with onsets, rather than with rimes, even in a visual task where articulation should be irrelevant.

Table 3. Results of mixed-effects logistic regression

Effect	β	SE	z	p
Intercept	-3.595	0.194	-18.536	< .001 *
O freq	0.075	0.017	4.441	< .001 *
M freq	0.078	0.016	5.019	< .001 *
R freq	0.221	0.015	14.278	< .001 *
O-M assoc	-0.312	0.016	-19.294	< .001 *
M-R assoc	-0.309	0.016	-19.054	< .001 *
O × M freq	0.150	0.015	9.845	< .001 *
M × R freq	-0.025	0.016	-1.535	.125
O × R freq	0.034	0.018	1.947	.052
O × M × R freq	0.013	0.017	0.761	.447

Notes: * represents statistical significance ($p < .05$). β, *SE*, *z*, *p* represent, respectively, the standardized regression coefficient, the standard error, the *z* value (computed using β and *SE*), and the *p* value (computed using *z*). O, M, and R represent onset, medial, and rime; assoc and freq stand for association and frequency.

But can this analysis be trusted? For example, we may ask whether it is valid to test syllables that are missing one or more syllable elements, since for these I arbitrarily set the element frequency to zero. There are 735 such items (out of the 3,274 total), so that removing all trials with them still leaves 78% of the data. Analysis of this subset gives virtually identical results: significant positive effects for all three element frequencies (βs > 0.2, ps < .0001), significant negative effects for both association strengths (onset-medial: β = -0.108, $p < .0001$; medial-rime: β = -0.058, $p < .0001$), a significant positive interaction between onset and medial frequencies (β = 0.047, p = .001), but no other interactions (βs < 0.02, ps > .2). Even the association strengths showed an asymmetry this time, with a significantly larger effect for the onset-medial grouping ($\chi^2(1)$ = 6.003, p = .01).

Could these asymmetries be due to some subtle confound among the three element frequencies, or to the inclusion of association strengths (which may overcomplicate the model), or to the lack of the standard (though here unnecessary) logarithm transformation on the frequencies? To address these questions, I ran another analysis using log-transformed element frequencies, with medial frequencies orthogonalized relative to onset and rime frequencies to eliminate confounding, and no association strengths. This time there was a significant positive interaction not just for onset frequency × medial frequency (β = 4.025, p = .009) but also medial × rime (β = 3.195, p = .03). As suggested by their standardized coefficients, however, these effects show the same asymmetry as before, with the onset-medial interaction significantly stronger ($\chi^2(1)$ = 10.564, p = .005). Thus this analysis merely reconfirms the previous ones.

5. Conclusions

Yip (2003) is right to complain that evidence regarding the position of the Mandarin medial is quite mixed. Yet this fact need not force us to embrace the null hypothesis that the medial has no intrinsic position at all. Theoretical phonologists, especially those who borrow data from neighboring fields like psycholinguistics or phonetics, should get into the habit of making explicit proposals about the nature of their data sources. In the case of the Mandarin medial, I believe that the disparate evidence across data sources falls out naturally from three competing factors: association strength (according to which the Mandarin medial shows no affiliation preference), articulation (where the medial is coarticulated with the onset), and

memory-constrained parsing (which favors associating the rime with the preceding medial).

I analyzed the results from an experiment where articulation and parsing were expected to play little role, namely a speeded acceptability judgment task with a very large number of written stimuli, in order to see which of the elements adjacent to the medial would interact more strongly with it, in the absence of these influences. I found that the medial consistently interacted more strongly with the onset. Thus unless it can be shown that articulation was somehow still relevant in this experiment (perhaps the participants subvocalized the test items before making their responses), or that yet another factor favoring onset-medial grouping came into play (perhaps the left-to-right visual display of the Zhuyin Fuhao symbols), it seems reasonable to agree, at least tentatively, with Duanmu (2007): the Mandarin medial is intrinsically grouped with the onset.

While I believe that my conclusion is based on more solid evidence than that adduced by Duanmu himself, I do not claim that my results are decisive. In addition to the above caveats, my experiment was not specifically designed to test the Mandarin medial question. I sincerely hope that other phonologists interested in this issue, whether they share my (tentative) conclusions or not, will critically examine my methods and propose their own. While doing so, however, I urge them to keep in mind the crucial role that methodological choices play in inference making, and not dismiss anomalous results as theoretically irrelevant or supportive only of the null hypothesis.

References

Aiken, Leona S., and Stephen G. West. *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications, 1991.

Anderson, John Mathieson, and Colin J. Ewen. *Principles of dependency phonology*. Cambridge: Cambridge University Press, 1987.

Baayen, R. Harald. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press, 2008.

Bailey, Todd M., and Ulrike Hahn. "Determinants of wordlikeness: Phonotactics or lexical neighborhoods?" *Journal of Memory and Language* 44, no. 4 (2001): 568-591.

Balota, David A., Melvin J. Yap, Keith A. Hutchison, and Michael J. Cortese. "Megastudies: What do millions (or so) of trials tell me about lexical processing?" In *Visual word recognition, Vol. 1: Models and methods, orthography and phonology*, edited by James S. Adelman, 90-115. Hove, England: Psychology Press, 2012.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. "Random effects structure for confirmatory hypothesis testing: Keep it maximal." *Journal of Memory and Language* 68, no. 3 (2013): 255-278.

Beckman, Mary, Jan Edwards, and Janet Fletcher. "Prosodic structure and tempo in a sonority model of articulatory dynamics." In *Papers in Laboratory Phonology II : Gesture, Segment, Prosody*, edited by Gerard J. Docherty and D. Robert Ladd, 68-86. Cambridge: Cambridge University Press, 1992.

Berg, Thomas, and Christian Koops. "The interplay of left-and right-branching effects: A phonotactic analysis of Korean syllable structure." *Lingua* 120, no. 1 (2010): 35-49.

Browman, Catherine P., and Louis Goldstein. "Some notes on syllable structure in articulatory phonology." *Phonetica* 45, no. 2-4 (1988): 140-155.

Chang, Yu-chu. "Phonotactic probability and the prenuclear glide in Mandarin Chinese." MA thesis, National Chung Cheng University, 2012.

Cheng, Chin-Chuan. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton, 1973.

Derwing, Bruce L., H. Samuel Wang, and Jane Tsay. "The syllable in Minnan: Making sense

of conflicting sources of evidence." In *Language and Cognition: Festschrift in Honor of James H-Y. Tai on His 70th Birthday*, edited by Jung-hsing Chang, 309-343. Taipei: The Crane Publishing, 2011.

Duanmu, San. *The phonology of Standard Chinese*, second edition. Oxford: Oxford University Press, 2007.

Duanmu, San. *Syllable structure: The limits of variation.* Oxford: Oxford University Press, 2010.

Hayes, Bruce, and Colin Wilson. "A maximum entropy model of phonotactics and phonotactic learning." *Linguistic Inquiry* 39, no. 3 (2008): 379-440.

Kapatsinski, Vsevolod. "Testing theories of linguistic constituency with configural learning: The case of the English syllable." *Language* 85, no. 2 (2009): 248-277.

Kimball, John. "Seven principles of surface structure parsing in natural language." *Cognition* 2, no. 1 (1973): 15-47.

Lee, Yongeun, and Matthew Goldrick. "The emergence of sub-syllabic representations." *Journal of Memory and Language* 59, no. 2 (2008): 155-168.

Levelt, Willem J. M. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press, 1989.

Li, Yingshing. "Investigating the status of the prenuclear glide in Mandarin Chinese using a maximum entropy model of phonotactic learning." Paper presented at National Conference on Linguistics, Yuan Ze University, Taiwan, November 27-28, 2009.

Marslen-Wilson, William, and Lorraine Komisarjevsky Tyler. "The temporal structure of spoken language understanding." *Cognition* 8, no. 1 (1980): 1-71.

McCarthy, John J. "Feature geometry and dependency: A review." *Phonetica* 45 (1988): 84-108.

Myers, James, and Jane Tsay. "The processing of phonological acceptability judgments." *Proceedings of Symposium on 90-92 NSC Projects,* 26-45. Taipei, 2005.

Schneider, Walter, Amy Eschmann, and Anthony Zuccolotto. *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc, 2002.

Tsai, Chih-Hao. "Mandarin syllable frequency counts for Chinese characters." Last modified 2000. http://technology.chtsai.org/syllable/.

Wan, I-Ping. "The status of prenuclear glides in Mandarin Chinese: Evidence from speech errors." *Chicago Linguistics Society* 33 (1997): 417-428.

Wang, H. Samuel, and Chih-ling Chang. "On the status of the prenucleus glide in Mandarin Chinese." *Language and Linguistics* 2, no. 2 (2001): 243-260.

Yip, Moira. "Casting doubt on the Onset-Rime distinction." *Lingua* 113, no. 8 (2003): 779-816.

Yoon, Yeo Bom, and Bruce Derwing. "A language without a rhyme: Syllable structure experiments in Korean." *Canadian Journal of Linguistics* 46 (2001): 187-237.