Versified Mandarin: Its Socio-Syntactic Parsing

The domain of "PF-Syntax" contains rich intra-speaker syntactic variation, made possible by spell-out operations like right dislocation. These variations could become associated with social meanings, such as style and prestige (Moore, 2023). In this study, I use PF constraints to explain a placement of BA-construction in Mandarin, which is ungrammatical except in verses. To further investigate the social significance of these PF constraints, I demonstrate how Mandarin verse syntax is parsed alongside metrical grids. This peculiarity reflects sociolinguistic ideology associated with verse-writing: metered syntax allows composers to write in an otherwise syntactically inaccessible style, performing prestige associated with features of archaic literary Chinese. Therefore, a socio-syntactic parser model can capture the process of analyzing and evaluating eccentric verse forms for virtuosity.

The Prosodic Syntax Theory of Chinese (Feng, 1995; 2017; 2019) identifies prosodic constraints that motivate syntactic movement repairs. A natural prediction of this theory is that prosody, when outranking syntax, could result in repairs that violates faithfulness expectations of the parser, to the point of ungrammaticality (Bertrand, 2023). Indeed, this is observed in verse lines. Most remarkably, the patient and the patient-marking BA^0 optionally precede the serial verb head *lái* in verses only.

(1) a. Mandarin: tā   lái   bǎ zuò yè   ná zǒu
                   She come   BA homework   pick up.

  b. Versified: [cp(tā   bǎ) | [np(zuò yè)] | [vp1 (lái [ná zǒuvp2])]]
                   She BA   homework   come pick up   (ungrammatical in prose)
          Both meaning: 'She comes to pick up the homework'

I explain this displacement using phonology, not syntax, especially since there is no semantic motivation to differentiate two structures. During versification, the string in (1)a violates an CONTINUITY constraint, as the word *zuò | yè* 'homework' is split by a foot boundary given default meter in a seven-syllable line (xx)(xx)|(xxx). Therefore, the metathesis of laí and bǎ zuò yè is a permissible rescue.

(2) ALIGN: Prosodic boundaries must overlaps with some syntactic phrasal boundary.
     CONTINUITY: No words (ω) can be split up by a prosodic boundary.

Another possible repair is to versify (1a) with a superfoot (bǎ zuò yè) in the middle. This, however, requires the whole stanza to fall in the metrical grid of (xx)(xxx)(xx), which is marked but still a possible meter. This reflects another crucial constraint:

(3) COHERE: Verses in the same group (i.e. a stanza) follow the same metrical grid.

According to Feng (2017), these constraints also exist in non-versified Mandarin, resulting in occasional morphosyntactic re-bracketing. Yet, versification frequently show evidence of metrical markedness constraints. This formal requirement compensates for the difficulty of parsing versified Mandarin. The genre of verse licenses elliptical and ambiguous strings: verse-writers frequently omit functional morphemes (classifier, copula, pronouns…), and re-linearize compounds. Therefore, much of parsing requires facilitation through metrical structure: verse-writers conventionally obey ALIGN, CONTINUITY and COHERE so that parsing is feasible in a highly economic metered textual medium. To demonstrate the utility of metrical parsing, I developed a toy parser through Python which uses metrical structure to bootstrap syntactic analysis of computational puzzles like Line 1.

(4)   Yīng tóu mǎ   shēn cháng yì   shòu   *Line 1*
           Eagle head horse body long wing beast
Parser Input: N   N   N   N   A   N   N
Parser Output: *#many possible parses* ([NN][NN][ANN], [NN][NNA][NN], etc.)
           Rén   ài   shòu shòu ài   yǔ zhòu   *Line 2*
           People love beast beast love universe

Parser Input: N   Vt   N   N   Vt   N
Parser Output: [[N   [Vt   N$_{VP}$]$_{CP}$][N [Vt   N$_{VP}$]$_{CP}$]
Meter Inference: (X   X   X) | (X   X) (X   X)

 Interpretation: 'People love (this) beast (, and) the beast loves the universe.'

 With only a few phrase structure rules, we can derive the unmarked structure of Line 2, yielding the metrical grid which separates words. Then, the parser extrapolate this rule to Line 1:

 (4) Eagle head horse   body long wing beast

Metrical Input: (X   X   X) | (X   X)   (X   X)
Parser Output: [N   N   N$_{NP}$] [N Ø$_{COP}$A CP] [N   N$_{NP}$]

 Interpretation: ~'Eagle-head horse / (and) that whose body is long — winged beast(s).'

 Rather than: ~'Eagle-head, horse-body / long-winged beast' (i.e. 'hippogriff') only available with the unmarked metrical grid (xx)(xx)|(xxx).

 On the sociolinguistic dimension, the prioritization of ALIGN and COHERE stems from the assumption that verses should not be parsed similar to speech processing. Instead, they ought to reflect prestigious archaisms of the classical language. Consequently, writers use unusual diction and structure in excess, making verses only comprehensible when chanted to an educated audience in the correct prosody. If text-setting violates metrical markedness constraints, they are ill-formed and incomprehensible. At the same time, faithfulness violation are still negative signs regarding the quality of writing. For this reason, verses like (1b) will be considered well-formed but awkward, lacking 'true virtuosity'.

 Translating constraint to sociolinguistic perceptions, I suggest a parser-comparator program that can analyze syntactic structure of Mandarin verses and predict reader evaluation of virtuosity. In its full form, it will take account of archaic/modern syntactic features, post-syntactic markedness, and lexical statistics to differentiate the forms of "virtuous" and "poorly-written" verses.

 Overall, linguistic judgment of verses is a novel and challenging problem at the intersection of formal and variationist linguistics. I use post-syntax metrical phonology to explain a Mandarin variation, which introduces the importance of a metrical parsing framework. This metrical parser system foregrounds the linguistic ideology of prestigious archaism which shapes grammatical and social perception of verses. Ongoing work develops a prototype of the socio-syntactic parser and experimentally investigate sociolinguistic factors that affect parsing and evaluation of Mandarin versification.

References

Bertrand, N., (2023) "When phonology outranks syntax: Postponed relative pronouns in Pindar", *Glossa: a journal of general linguistics* 8(1). doi: https://doi.org/10.16995/glossa.9842

Feng, S. (1995). *Prosodic structure and prosodically constrained syntax in Chinese* [Doctoral Dissertation, University of Pennsylvania]. https://www.proquest.com/docview/304205447?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Feng, S. (2017). *Prosodic Morphology in Mandarin Chinese*. New York, NY. Routledge https://doi.org/10.4324/9781315392783

Feng, S. (2019). *Prosodic Syntax in Chinese: History and Changes*. New York, NY. Routledge. https://doi.org/10.4324/9781351263245

Hayes, B., & Kaun, A. (1996). "The role of phonological phrasing in sung and chanted verse", *The Linguistics Review. 13*(3–4), 243–304. https://doi.org/10.1515/tlir.1996.13.3-4.243

McCarthy, J. J., & Prince, A. (2004). Generalized Alignment: Prosody. In *Optimality Theory in Phonology* (pp. 165–177). Hoboken, NJ. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756171.ch7

Moore, E. (2023). *Socio-syntax: Exploring the Social Life of Grammar*. Cambridge, UK. Cambridge University Press. https://doi.org/10.1017/9781108921299