

Syntactic patterns in classical Chinese poems: A quantitative study

John Lee, Yin Hei Kong, Mengqi Luo
2017

Подготовила: Анна
Дмитриева, МКЛ171

Задачи исследования

- Квантитативный анализ параллелизма в поэзии династии Тан. Затрагиваются такие особенности, как:
 - Синтаксический параллелизм;
 - Образность и пропозициональность языка.
- Данные:
 - Complete Tang Poems (всего 2000 стихов, деревья разбора сделаны для 971, 617 составляют датасет).

Предмет исследования

- Четырёхкуплетная поэзия с 起-承-转-合 (qǐ-chéng-zhuǎn-he) структурой:
 - Первый куплет (起): задаёт время и место;
 - Второй (承): развивает сюжет при помощи параллельных образов;
 - Третий (转): «сюжетный поворот» при помощи противоположных образов;
 - Четвёртый (合): заключение, часто субъективное.

杜甫, «江汉»/Ду Фу, «Между Янцзыцзяном и рекой Хань» (пер. А.И. Гитовича)

江汉思归客，乾坤一腐儒。
片云天共远，永夜月同孤。
落日心犹壮，秋风病欲苏。
古来存老马，不必取长途。

jiāng hàn sī guī kè, gān kūn
yī fǔ rú.

piàn yún tiān gòng yuǎn,
yǒng yè yuè tóng gū.

luò rì xīn yóu zhuàng, qiū
fēng bìng yù sū.

gǔ lái cún lǎo mǎ, bù bì qǔ
zhǎng tú.

Я - путник, скитающийся давно
Меж двух величавых рек,
Ненужный ученый - в чужом краю
Затерянный человек.
Брожу я от родины вдалеке,
И некому мне помочь,
И я одинок, подобно луне
В долгую зимнюю ночь.
Близится горестный мой закат,
Но душа еще молода.
Быть может, не будут болезни мои
Мучить меня всегда?
Я слышал, что в древние времена
Кормили старых коней
Отнюдь не за то, что они могли
Работать на склоне дней.

Параллелизм

泉	声	咽	危	石
Quán	shēng	yàn	wēi	shí
fountain	sound	gurgles	precipitous	rock
f	N	V	f	N
The fountain sound gurgles above precipitous rocks				
日	色	冷	青	松
rì	sè	lěng	qīng	sōng
sun	color	cold	blue	pine
f	N	F	f	N
Sun rays are chilly amidst the blue pines				

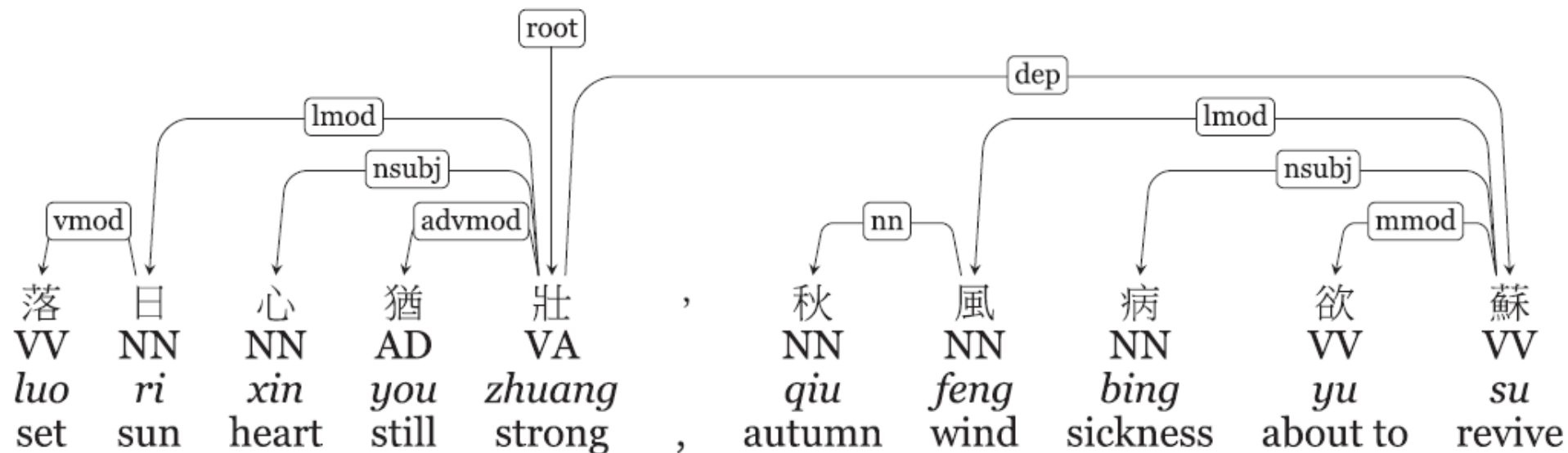
Строки (части строки) должны иметь одинаковую структуру, слова на одних и тех же позициях должны принадлежать одинаковым частям речи и иметь сходное значение.

Образность и пропозициональность

- Образный язык (imagistic language): изображения и чувства. Близок к языку поэзии имажинистов;
 - Преобладают именные группы, связь между фразами явно не выражена:
 - 江汉思归客(N), 乾坤一腐儒(N): две именные группы
- Пропозициональный язык: утверждения
 - Больше функциональных слов, синтаксис действия
 - 古来存老马(N, subj), 不(Adv, neg)必(V modal)取(V, main)长途
- Традиционно вторая и третья строка образные, четвёртая – пропозициональная / строки в середине заканчиваются пропозициональными утверждениями.

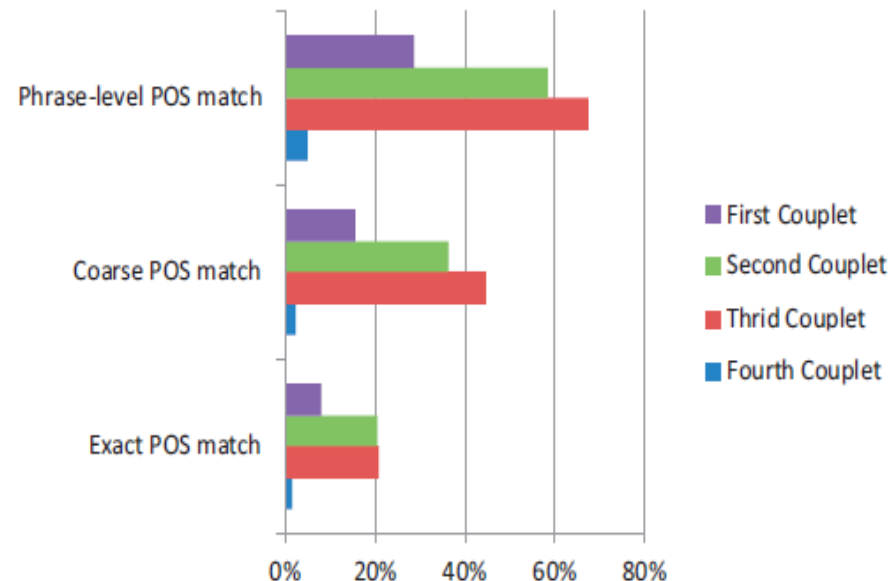
POS-теги и деревья разбора

- Теги и деревья разбора взяты из Penn Chinese Treebank;
- Убраны теги только для современного китайского, добавлены связи, характерные для классического языка;
- Теги присваиваются каждому иероглифу: например, слово 黄河/Хуанхэ получает теги 黄(Adj)河(Noun) вместо 黄河 (Pronoun)



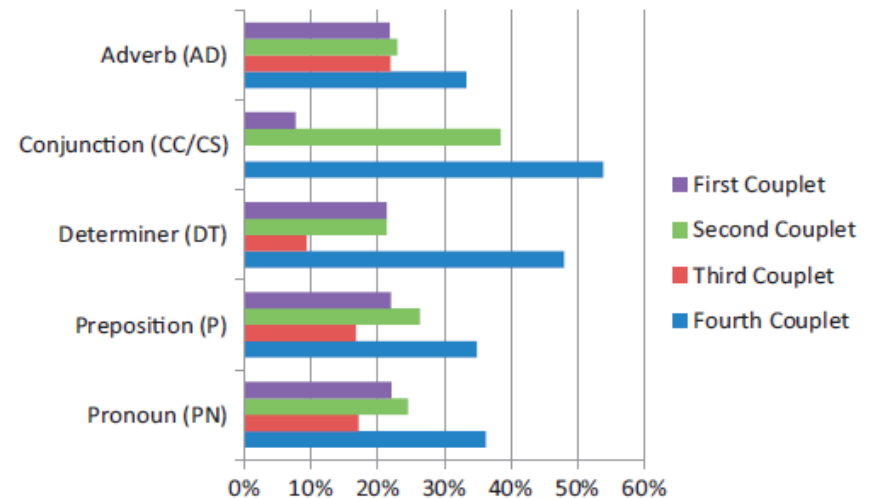
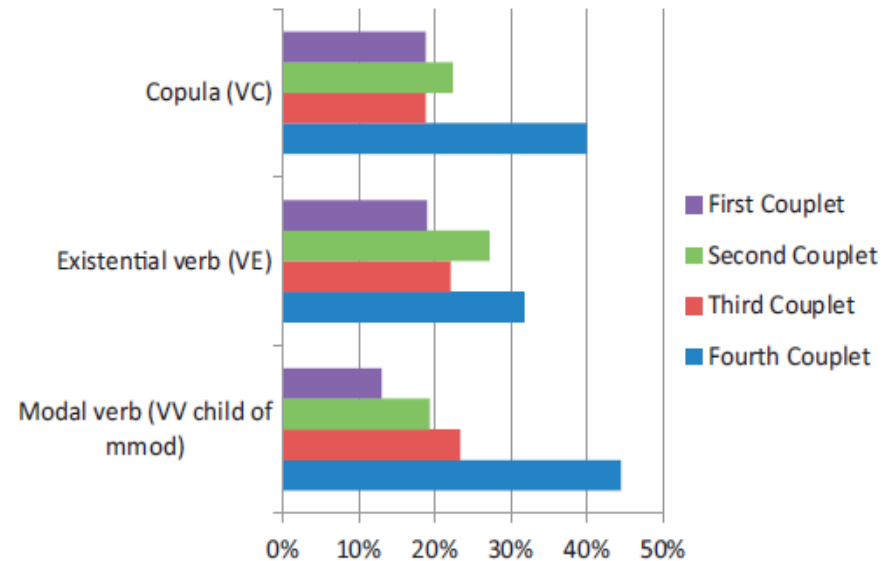
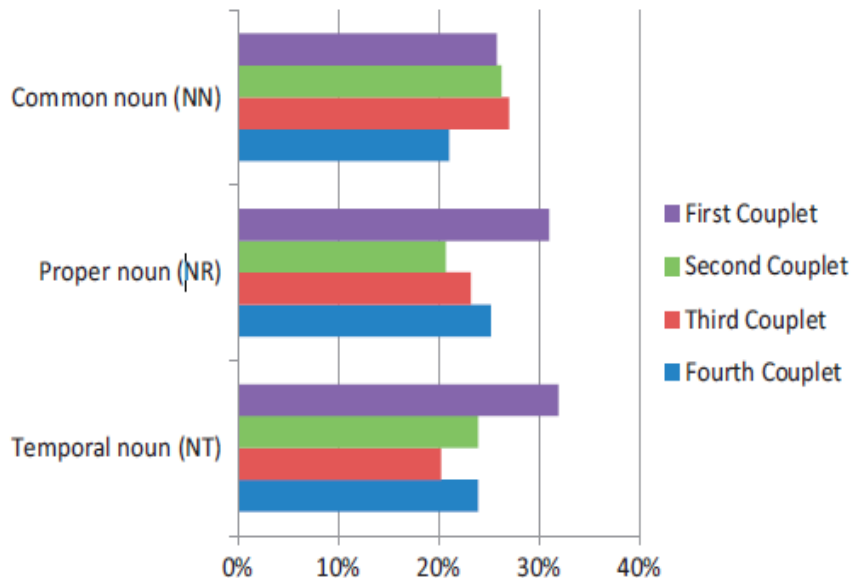
Анализ параллелизма

- POS-теги на уровне слов: каждый иероглиф в первой части строки соответствует по части речи иероглифу на той же позиции во второй части строки;
- Проблема: слишком подробные теги (NN, NR, NT, PN...). Решение: объединение тегов (например, VA, VC, VE, VV => V*);
- POS-теги на уровне фраз: теги параллельны в контексте фразы. Например, в NP модификаторами могут быть как прилагательные, так и существительные, и такие группы будут считаться параллельными



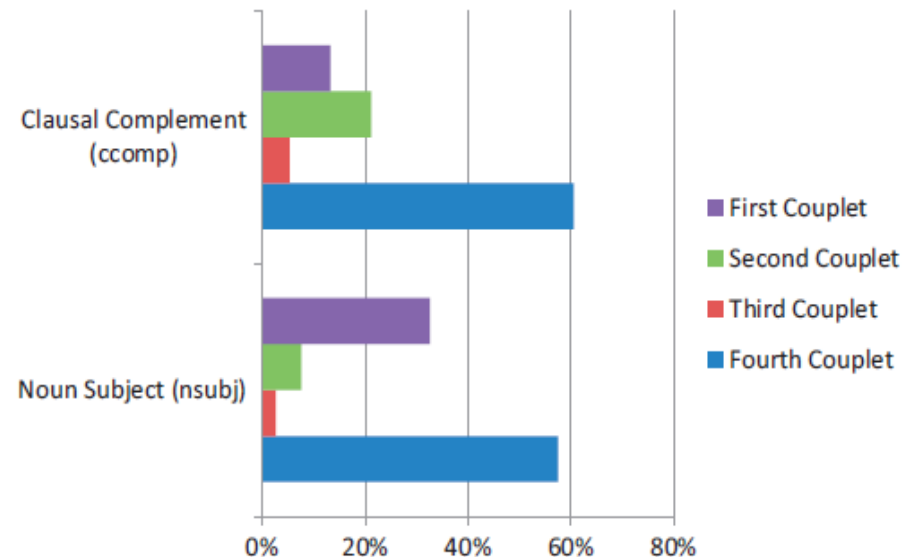
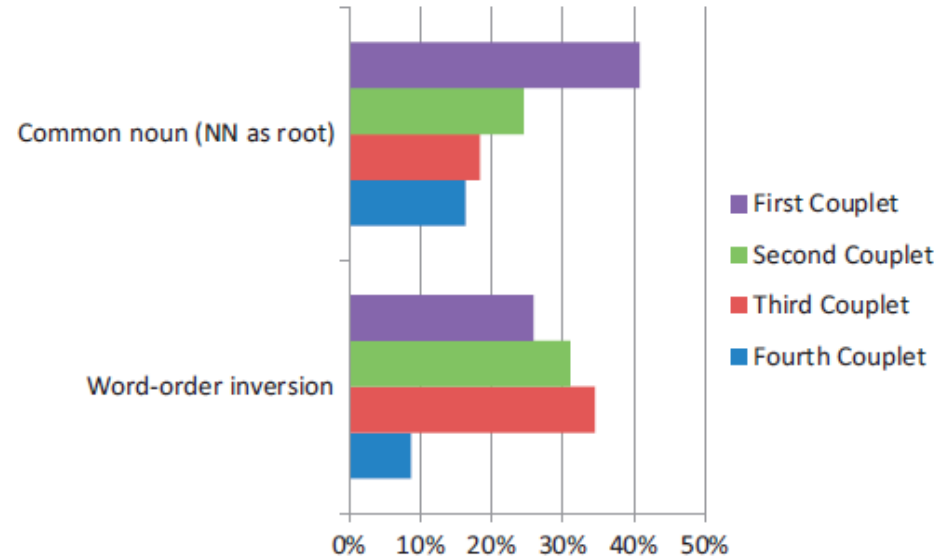
Анализ параллелизма: образность

- Процент имен и функциональных слов



Анализ параллелизма: образность

- Фрагментарность синтаксиса: процент куплетов, состоящих только из именных групп, и процент изменения порядка слов;
- Синтаксическая целостность: отношения зависимости объединяют две строки в куплете.



Выводы

- Большинство «новых стихов» действительно имеют непараллельный четвёртый куплет; чуть более половины всех стихов в выборке имели параллельный первый куплет и непараллельный второй или третий (偷春);
- Первый, второй и третий куплеты обычно образны;
- Четвёртый куплет в целом более пропозиционален, менее декларативен и имеет более связную структуру;
- Чем меньше групп тегов, тем больше параллелизма.