

МОДЕЛИРОВАНИЕ РИТМИКИ СТИХА ПОСРЕДСТВОМ МАРКОВСКИХ ПРОЦЕССОВ

ЯНА КЛИМЕНТОВА (Прага)

1. Введение

Статистические исследования стиха, проводимые лингвистами и математиками, не представляют уже продолжительное время никаких новостей. Они делают возможным доказательство индивидуальных предпосылок, проверяют правильность или неправильность гипотез.

Статистические методы используются при исследовании разных характеристик стиха, как напр. рифмы, ритмики итд. К исследованию ритмики стиха возможно подойти в основном двумя способами:

- а) стихи исследуются вертикально,
- б) стихи исследуются горизонтально.

Первый метод имеет свою традицию. Он применялся Матезиусом, Мука-ржовским и Томашевским. Как показывает Левы в одной из своих последних работ,¹ этот метод удобен для чешского языка. Но если нашей целью является моделирование чередования слогов, несущих метрический импульс, со слогами лишенными этого импульса посредством Марковского процесса, оказывается более удобным второй метод, который в последнее время предпочитается. Однако, исследование определенных зависимостей в этом процессе требует тоже использование первого метода. В конце своего доклада попытаюсь доказать возможность применения вертикального способа исследования стиха для моделирования ритмики стиха.

Статья является примером исследования ритмики стиха посредством аппарата стохастических процессов. Приводимые здесь результаты я не буду обобщать, хотя в будущем я бы хотела расширить применяемый метод на исследование творчества большего количества поэтов и сравнить стохастические процессы в разных произведениях.

2. Постановка проблемы

В своей статье я буду заниматься моделированием процесса чередования слогов несущих метрический импульс (в тоническом стихе речь идет о ударении в слове) со слогами характеризованными отсутствием этого качества. Анализироваться будут два стихотворения.

¹ Jiří Levý, *Předběžné poznámky k informační analýze verše*, Slovenská literatura XI, 1964, стр. 15.

Слог, на который падает ударение в слове, я буду обозначать S_1 и слог без ударения S_2 .

В приведенном процессе я бы хотела исследовать зависимость появления S_1 или S_2 на определенной позиции в стихе от предыдущих позиций и показать, может ли процесс быть гомогенным и стационарным.

3. Выбор материала и определение основных терминов

Процесс чередования S_1 и S_2 я исследовала в четырехстопном ямбе и хоре в произведении В. Галка „В природе“ и в произведении Я. Врхлицкого „Сатанела“. Далее буду обозначать эти произведения „совокупность 1“ и „совокупность 2“.

При определении явлений S_1 и S_2 я старалась избежать влияния ритма и не принимать во внимание побочные ударения.² Необходимо признать, что это является определенным упрощением проблемы различения ударных и неударных слогов в стихе. В статистических исследованиях необходимо прежде всего соблюдать установленный метод, чтобы результаты могли дальше сравниваться.

В качестве модели чередования S_1 и S_2 можно себе представить систему с двумя состояниями и столькими временами, сколько слогов в стихе. Одним состоянием будет явление S_1 и вторым состоянием явление S_2 .

В данной статье используется термин вероятность, хотя в действительности принимается во внимание относительная частота, то есть оценка вероятности.

Вероятность появления состояния S_1 или S_2 на i -той позиции мы обозначим $p_1^{(i)}$ или $p_2^{(i)}$, где $i = 1, 2, \dots, 11$. i -тая позиция определяется как i -тый порядок слога в стихе. Индекс i может быть величины $1, 2, \dots, 8$, потому что были выбраны произведения, написанные четырехстопным стихом. Однако, в вводных стихах совокупности 1, которые случайно попали в выборку, обнаружилось еще 11 позиций, которые для полноты получаемых результатов было необходимо учесть, то есть $i = 1, 2, \dots, 11$.

Условную вероятность того, что система находится в состоянии S_1 на i -той позиции, с предпосылкой, что на $(i-1)$ -ой позиции была в состоянии S_2 , я обозначу $p_{21}^{(i,1)}$. Вероятность $p_{21}^{(i,1)}$ называется вероятностью перехода от состояния S_2 к состоянию S_1 по одному шагу. Также обозначается вероятность перехода от S_1 к S_2 по одному шагу $p_{12}^{(i,1)}$.

Дальнейшим типом вероятности, который приводится здесь, является вероятность останова в состоянии S_1 по одному шагу, обозначенная $p_{11}^{(i,1)}$. Это вероятность появления S_1 непосредственно после S_1 . Вероятность останова в состоянии S_2 , т. е. вероятность того, что непосредственно после S_2 приходит S_2 , я обозначу $p_{22}^{(i,1)}$.

Определенные здесь вероятности я исследовала отдельно в совокупности 1 и в совокупности 2; они содержатся в таблице 1 для совокупности 1 и в таблице 2 для совокупности 2.

Далее, я вычислила условные вероятности появления состояния S_1 (S_2) при условии, что оно явится непосредственно после двух определенных состояний. Я вычислила, например, вероятность того, что если $(i-2)$ -ый слог

² Josef Hrabák, *Úvod do teorie verše*, стр. 36.

стиха несет ударение, а $(i-1)$ -ый слог неударяем, то i -тый слог в стихе будет ударный. Эту вероятность обозначу $p_{121}^{(i,1)}$. Дальше я определила вероятности $p_{211}^{(i,2)}$, $p_{221}^{(i,2)}$, $p_{111}^{(i,2)}$, $p_{122}^{(i,2)}$, $p_{212}^{(i,2)}$, $p_{222}^{(i,2)}$, $p_{112}^{(i,2)}$, значение которых можно определить подобным способом, как я показала выше у вероятности $p_{121}^{(i,2)}$. Эти вероятности приведены в таблице 3 для совокупности 1 и в таблице 4 для совокупности 2.

Пример:

В таблице 3 найдется вероятность перехода $p_{121}^{(1,2)}$ то есть вероятность того, что на третьей позиции стиха будет S_1 , если на второй позиции было S_2 и на 1-ой S_1 , в строчке обозначенной S_1S_2 и в столбике S_1 в позиции 123, значит, в правой части таблицы в последней строчке найдется под номерами 123, что $p_{121}^{(1,2)} = 0,54$.

В таблицах 1, 2, 3, 4 находятся матрицы вероятностей перехода по одному шагу, так наз. стохастические матрицы. Например, из таблицы 1 для вероятностей перехода $p_{kl}^{(1,1)}$ получим матрицу

$$\begin{pmatrix} p_{22}^{(1,1)} & p_{21}^{(1,1)} \\ p_{12}^{(1,1)} & p_{11}^{(1,1)} \end{pmatrix} = \begin{pmatrix} 0,04 & 0,96 \\ 0,97 & 0,03 \end{pmatrix},$$

которую я буду дальше обозначать просто $(p_{kl}^{(1,1)})$ и обобщенно $(p_{kl}^{(i,1)})$, где $k, l = 1, 2; i = 1, 2, \dots, 11$.

Таким же образом можно напр. из таблицы 3 получить стохастическую матрицу для вероятностей перехода $p_{klm}^{(1,2)}$

$$\begin{pmatrix} p_{222}^{(1,2)} & p_{221}^{(1,2)} \\ p_{212}^{(1,2)} & p_{211}^{(1,2)} \\ p_{112}^{(1,2)} & p_{111}^{(1,2)} \\ p_{122}^{(1,2)} & p_{121}^{(1,2)} \end{pmatrix} = \begin{pmatrix} 0,13 & 0,87 \\ 0,99 & 0,01 \\ 1,00 & 0,00 \\ 0,46 & 0,54 \end{pmatrix},$$

которую я буду обозначать $(p_{klm}^{(1,2)})$ и обобщенно для i -той позиции введу обозначение стохастической матрицы $(p_{klm}^{(i,2)})$.

4. Теоретическая модель ямба и хорей

Если бы в стихотворениях строго соблюдался размер (то есть четырехстопный ямб или хорей), было бы возможно стохастические матрицы $(p_{kl}^{(i,1)})$ заменить единственной матрицей независимой от позиции в стихе и обозначаемой $(p_{kl}^{(1)})$,

$$(p_{kl}^{(1)}) = \begin{pmatrix} p_{22}^{(1)} & p_{21}^{(1)} \\ p_{12}^{(1)} & p_{11}^{(1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Также можно, в случае теоретической модели, стохастические матрицы $(p_{klm}^{(i,2)})$ заменить единственной матрицей независимой от позиции в стихе, которую я обозначу

$$(p_{klm}^{(2)}) = \begin{pmatrix} p_{222}^{(2)} & p_{221}^{(2)} \\ p_{212}^{(2)} & p_{211}^{(2)} \\ p_{112}^{(2)} & p_{111}^{(2)} \\ p_{122}^{(2)} & p_{121}^{(2)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Теоретический процесс, определяемый абсолютными вероятностями появления S_1 , которые обозначу p_1 , и появление S_2 , которые обозначу p_2 , и стохастической матрицей $(p_{kl}^{(1)})$ называется простым Марковским процессом, гомогенным и стационарным.

Процесс является простым Марковским потому, что вероятность появления состояния S_1 и S_2 зависит только от состояния непосредственного предыдущего, то есть, имеют силу следующие уравнения

$$p_{klm}^{(2)} = p_{lm}^{(1)}$$

для $k, l, m = 1, 2$. Сила этих уравнений вытекает из стохастических матриц $(p_{klm}^{(2)})$ и $(p_{lm}^{(1)})$.

Гомогенность теоретического процесса вытекает из независимости вероятности перехода от позиции в стихе, то есть от индекса i .

Теоретический процесс стационарен по той причине, что после двух слогов состояния повторяются.

Приведенные свойства теоретической модели нельзя обобщить в отношении к реальному ритмическому процессу. Колмогоров и Кондратов в своей работе о ритмике Маяковского отмечают, что нельзя во что бы то ни стало искать закономерность или регулярность, но что надо также учитывать существование неоднородных частей. Дальше я попытаюсь показать и оправдать отклонения реального ритмического процесса от его теоретической модели и покажу, каким образом становится возможным моделировать ритмику стиха Марковскими процессами.

5. Реальный ритмический процесс

5.1 Память реального ритмического процесса

Прежде чем заниматься дальнейшими свойствами реального процесса, необходимо задуматься над тем, какой памятью исследуемый процесс обладает. Память — это зависимость появления состояния системы от появления состояний на предыдущих позициях. Если процесс полностью определяется вероятностями перехода $p_{kl}^{(i,1)}$, то это процесс с простой памятью. Если процесс определяется вероятностями перехода $p_{klm}^{(i,2)}$, то это процесс с двойной памятью. Если появление состояния на определенной позиции зависит от появления состояний на нескольких предыдущих позициях, то мы говорим о процессе с многократной памятью.

Таким процессом с многократной памятью, по-моему, является тоже исследуемый процесс. Во всяком случае трудно определить размер памяти. Этот размер определяется количеством позиций, от которых зависит состояние на исследуемой позиции. Считаю, что количество позиций, от которых зависит появление S_1 и S_2 на определенной позиции, испытывает влияние (кроме зависимостей на предыдущих позициях) тоже

- а) длины слов в стихе,
- б) предыдущих стихов,
- в) длины предложений в стихотворении.

5.1.1. Зависимость от предыдущих позиций

Сначала я буду принимать во внимание только зависимость появления состояний S_1 и S_2 на двух непосредственно предшествующих позициях, не смотря на влияния а), б), в). Очевидно, что появление состояния S_1 или S_2 на i -той позиции стиха зависит от состояния, которое появится на $(i-1)$ -ой позиции. В противном случае не было бы процесса вообще. Я хотела определить, зависит ли или не зависит появление исследуемых состояний на i -той позиции от появления состояний на $(i-1)$ -ой позиции в стихе. Эту зависимость я проверяла помощью критерия χ^2 . Результаты критерия показали значительную разницу между совокупностями 1 и 2. Для совокупности 2 я определила, что появление S_1 (S_2) на i -той позиции стиха не зависит от $(i-2)$ -ой позиции, между тем как в совокупности 1 эта зависимость большая. Поэтому можно судить на регулярность стиха Врхлицкого.

По приведенному критерию можно судить, что совокупность 2 представляет собой процесс с простой памятью. Однако, если положение на i -той позиции не зависит от $(i-2)$ -ой позиции, это не значит, что оно не зависит тоже от $(i-3)$ -ей позиции в стихе. Зависимость i -той позиции от $(i-3)$ -ей позиции стиха пока критерию не подвергалась. На основе этих рассуждений о размере памяти процесса нельзя сказать ничего.

5.1.2 Зависимость от длины слов³

Длина слова влияет на появление S_1 или S_2 на каждой позиции по разному. Это явление хотела бы я иллюстрировать на примере первых четырех слогов четырехстопного ямба, для которого я получила результаты на основе совокупности 1.

На первой позиции этого размера S_2 явится с вероятностью 0,77, на второй позиции с вероятностью 0,75 будет S_1 и на третьей позиции с вероятностью 0,85 явится S_2 . Вероятности $p_2^{(1)}$, $p_1^{(2)}$, $p_2^{(3)}$ близки вероятностям приводимым для теоретической модели. Это можно доказать стремлением поэта соблюдать ритм в начале стиха. Этот факт подтверждается тоже результатами приведенными в таблице 1. Стохастическая матрица для вероятности перехода с первой позиции на вторую.

$$(p_{ki}^{(1,1)}) = \begin{pmatrix} 0,04 & 0,96 \\ 0,97 & 0,03 \end{pmatrix}$$

приближается матрице теоретической модели

$$(p_{ki}^{(1)}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Появление S_1 или S_2 на четвертой позиции стиха зависит от того, имеется ли на второй позиции стиха слово двухслоговое или многослоговое. Так как количество многослоговых слов 41,3 %, между тем как двухслоговых только 22,5 %, то вероятность того, что на второй позиции будет начинать слово многослоговое, будет больше, чем вероятность того, что на второй позиции будет начинать слово двухслоговое, а также вероятность появления

³ Jiří Levý, *Matematický a experimentální rozbor verše*, Česká literatura 12, 1964, стр. 218.

Таблица 1

Вероятн остиперехода $p_{ii}^{(1)}$ для совокупности 1

позиция	S_2									
	12	23	34	45	56	67	78	89	910	1011
S_1	0,04	0,43	0,28	0,48	0,12	0,63	0,52	0,98	0,57	1,00
S_2	0,97	0,99	1,00	0,98	0,99	1,00	1,00	1,00	0,00	1,00
позиция	S_1									
	12	23	34	45	56	67	78	89	910	1011
S_1	0,96	0,57	0,72	0,52	0,88	0,37	0,48	0,02	0,43	0,00
S_2	0,03	0,01	0,00	0,02	0,01	0,00	0,00	0,00	0,00	0,00

Таблица 2

Вероятности перехода $p_{ii}^{(1)}$ для совокупности 2

позиция	S_2							S_1						
	12	23	34	45	56	67	78	12	23	34	45	56	67	78
S_1	0,36	0,39	0,98	0,02	0,46	0,37	1,00	0,64	0,61	0,02	0,98	0,54	0,63	0,00
S_2	1,00	0,86	1,00	0,80	0,99	0,67	1,00	0,00	0,14	0,00	0,20	0,01	0,33	0,00

Таблица 3

Вероятности перехода $p_{ii}^{(2)}$ для совокупности 1

позиция	S_2									
	123	234	345	456	567	678	789	8910	91011	
S_2S_2	0,13	0,35	0,71	0,04	0,79	0,31	1,00	0,00	1,00	
S_2S_1	0,99	1,00	0,98	1,00	1,00	1,00	1,00	0,00	1,00	
S_1S_1	1,00	1,00	0,00	0,80	1,00	0,00	0,00	0,00	0,00	
S_1S_2	0,46	0,28	0,09	0,14	0,57	0,56	0,94	0,69	0,00	
позиция	S_1									
	123	234	345	456	567	678	789	8910	91011	
S_2S_2	0,87	0,65	0,29	0,96	0,21	0,69	0,00	1,00	0,00	
S_2S_1	0,01	0,00	0,02	0,00	0,00	0,00	0,00	1,00	0,00	
S_1S_1	0,00	0,00	0,00	0,20	0,00	0,00	0,00	0,00	0,00	
S_1S_2	0,54	0,72	0,91	0,86	0,43	0,44	0,06	0,31	0,00	

Таблица 4
Вероятности перехода $p_{iim}^{(i,2)}$ для совокупности 2

позиция	S_2						S_1					
	123	234	345	456	567	678	123	234	345	456	567	678
S_2S_2	0,00	0,97	0,02	0,33	0,00	1,00	1,00	0,03	0,98	0,67	1,00	0,00
S_2S_1	0,86	1,00	0,80	0,99	0,57	1,00	0,14	0,00	0,20	0,01	0,43	0,00
S_1S_1	0,00	1,00	0,00	1,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
S_1S_2	0,40	1,00	0,02	0,75	0,38	1,00	0,60	0,00	0,98	0,25	0,62	0,00

S_1 на четвертой позиции ($p_1^{(4)} = 0,6$) будет незначительна в сравнении с предполагаемой вероятностью ($p_1 = 1$).

Таким же образом можно поступать при исследовании дальнейших позиций в стихе. Из рассуждений, приведенных в этой части, вытекает, что процесс имеет размер памяти равный количеству слогов самого длинного слова в стихе.

5.1.3 Зависимость от предшествующих стихов

Ситуация в конце стиха может зависеть от групп S_1 и S_2 в конце предшествующих стихов. Эту зависимость нельзя анализировать рассматриваемым процессом, который относится только к одному стиху. Оказывается, что будет удобно для этой цели моделировать ритмику стиха другой моделью, о которой я буду говорить дальше.

5.1.4 Зависимость от длины предложений

Однословные слова, которые представляют 26,4 % чешской лексики,⁴ проблематичны с точки зрения ударения. Однословные слова в определенных ситуациях являются носителями метрического импульса. Такие ситуации могут испытывать влияние ударения предложения.

Предложение обыкновенно превосходит длину одного стиха, следовательно зависимость появления S_1 или S_2 от длины предложения нельзя выразить помощью рассматриваемого процесса, который может дать анализ состояния внутри стиха, но не показывает отношение между стихами.

5.2 Гомогенность реального ритмического процесса

Гомогенность процесса определяется в 4-ой части. Если процесс считать гомогенным, должны иметь силу уравнения

$$p_{12}^{(i,1)} = p_{12}^{(i+1,1)}, p_{22}^{(i,1)} = p_{22}^{(i+1,1)}, p_{21}^{(i,1)} = p_{21}^{(i+1,1)}, p_{11}^{(i,1)} = p_{11}^{(i+1,1)},$$

это значит, что в таблице 1 и 2 должны быть одинаковые номера в 1-ой строчке левой части, в первой строчке правой части, во второй строчке левой части и во второй строчке правой части.

Из таблиц 1 и 2 ясно, что приведенные равенства не входят в силу, и, следовательно реальный процесс не гомогенен.

⁴ См. там же.

5.3 Стационарность реального ритмического процесса

На основе сравнительно небольшой выборки, которую я проводила, видны определенные признаки стационарности. После четырех шагов вероятности появления пар состояний похожи. Вероятности появления пар состояний я обозначу $p_{ki}^{(i,i+1)}$. Напр., вероятность появления пары S_1S_2 на второй и третьей позиции обозначу $p_{12}^{(2,3)}$.

В совокупности 1 вероятности ($p_{ki}^{(1,2)}$) (0,04; 0,74; 0,01; 0,21) близки вероятностям $p_{ki}^{(5,6)}$ (0,09; 0,69; 0,01; 0,21). Такие результаты получены для вероятностей $p_{ki}^{(2,3)}$, $p_{ki}^{(6,7)}$, и $p_{ki}^{(3,4)}$, $p_{ki}^{(7,8)}$. Отклонения в конце стиха уже велики. Они возникли тем, что только в незначительном количестве стихов было больше чем 8 позиций и поэтому вероятности появления пар состояний не определяются с такой точностью, как вероятности пар в середине стиха. Такие же заключения действительны в совокупности 2.

К более точным результатам возможно прийти исследованием совокупностей большего размера. Полученные мною результаты подсказывают, что ритмический реальный процесс для рассматриваемых стихотворений стационарен и что явления повторяются после четырех шагов.

6. Заключение

На основе проведенных рассуждений можно сказать о процессе чередования S_1 и S_2 в совокупности 1 и 2, что это процесс с памятью объема n , где n больше чем 4, при условии, что в стихотворении появится хоть одно четырехсловное слово. В современном состоянии исследования верхняя граница для n неизвестна. Процесс не является гомогенным и оказывается, что он стационарен.

7. Другая модель ритмики стиха

Ритмику стиха можно моделировать тоже другим способом. Покажу на примере четырехстопного ямба.

В основе модели лежит определение отклонений от стиха предписанного размером $\cup - \cup - \cup - \cup -$. Например, отклонение $\cup - \cup\cup \cup - \cup\cup$ было бы обозначено а, отклонение $\cup\cup \cup - \cup - \cup -$ б итд. Состояниями системы были бы типы отклонений а, б, ... Параметром было бы число стиха в стихотворении. В таком процессе были бы установлены зависимости между стихами, о которых выше приведенный процесс ничего не говорит. Дальше появились бы отклонения от теоретической модели.

MODELOVÁNÍ RYTMU VERŠE S POMOCÍ MARKOVOVSKÝCH PROCESŮ

Práce je ukázkou, jak lze postupovat při zkoumání rytmické výstavby verše pomocí aparátu stochastických procesů. Je zde modelován proces střídání slabik, na které připadá prvek nesoucí metrický impuls, se slabikami charakterizovanými nedostatkem tohoto prvku. Slabika, která je nositelkou slovního přízvuku, je označována S_1 a slabika postrádající tento prvek S_2 .

V procesu střídání S_1 a S_2 se vyšetřuje závislost výskytu S_1 nebo S_2 na určité pozici ve verši na předchozích pozicích, na délce slov ve verši, na délce vět v básni a na předchozích verších, tzv. paměť stochastického procesu. Dále se zjišťuje, zda proces je homogenní a stacionární.

Uvedený proces se zkoumá v čtyřstopém jambu a trocheji v díle V. Hálek, V přírodě a v díle J. Vrchlického, Satanela.

MODELLING THE RHYTHM OF THE LINE BY MEANS OF THE MARKOV PROCESSES

This paper shows how to proceed in the investigation of the rhythmic structure of the line by using the stochastic processes apparatus. We have here modelled the process of variation of the syllables on which the element bearing the metrical impulse falls, with the syllables characterized by the lack of this element. The syllable which bears the word stress is designated S_1 and the syllable without this element S_2 .

In the process of alternating S_1 and S_2 , the author examines to what extent the appearance of S_1 or S_2 in a certain position in the line depends on earlier positions, on the length of words in the line, on the length of sentences in the poem and on preceding lines, the so-called memory of the stochastic process. Further investigation is carried out as to whether the process is homogeneous and stationary.

The process mentioned is examined in the four-foot iamb and trochee in V. Hálek's work V přírodě and in J. Vrchlický's Satanela.

Translated by Jessie Kocmanová

