

Евразийский гуманитарный журнал. 2025. № 1. С. 21–29.

Eurasian Humanitarian Journal. 2025. No. 1. P. 21-29.

Научная статья

УДК 811.581(07)

МОДЕЛЬ ОЦЕНКИ СЛОЖНОСТИ ЯЗЫКОВ: КИТАЙСКИЙ ЯЗЫК С ТОЧКИ ЗРЕНИЯ АНГЛИЙСКОГО ЯЗЫКА

Андрей Андреевич Осадчий¹, Светлана Викторовна Андросова²

^{1,2}Амурский государственный университет, Благовещенск, Россия,

osadchiiandrei2001@gmail.com

Аннотация. Данная статья посвящена попытке формализации существующей модели оценки сложности английского языка. Модель основана на типологическом подходе классификации языков, выработанном на опыте Отраслевой лаборатории машинного перевода Российского государственного педагогического университета им. А. И. Герцена. Цель настоящей статьи – предложить авторское расширение и формализацию модели оценки сложности языков по системе, разработанной Foreign Language Institute (FSI), для оценки сложности любого иностранного языка с точки зрения английского языка, и апробировать ее на известных данных по характеристикам китайского языка (фонетика, лексика, грамматика, система письма и др.). Основным методом исследования является моделирование. Автором статьи разбирается устройство данной модели, её критерии (переменные), изменение её результатов, и предлагается математическая формализация данной модели с последующими вычислительными экспериментами. В результате для китайского языка были рассчитаны два параметра: значение структурной сложности, которое составило 69 %, и индекс FSI, который оказался приблизительно равным 4,573. Эти значения позволили присвоить китайскому языку категорию V. Данная категория означает, что среднестатистическому человеку, изучающему китайский язык, для которого родным является английский, потребуется 85 недель для достижения уровня B2/C1.

Ключевые слова: английский язык, сложности изучения языка, математическая модель, формализация, уровень языка, языковое родство, письменность.

Для цитирования: Осадчий А. А., Андросова С. В. Модель оценки сложности языков: китайский язык с точки зрения английского языка // Евразийский гуманитарный журнал 2025. № 1. С. 21–29.

Original article

THE MODEL TO ASSESS THE DIFFICULTY OF A LANGUAGE: CHINESE FROM THE VIEW POINT OF ENGLISH

Andrei A. Osadchi¹, Svetlana V. Androsova²

^{1,2}Amur State University, Blagoveshchensk, Russia, osadchiiandrei2001@gmail.com

Abstract. This article is devoted to an attempt to formalize the existing model for assessing the complexity of the English language. The model is based on a typological approach to language classification, developed from the experience of the Industrial Laboratory of Machine Translation of the Russian State Pedagogical University named after A. I. Herzen. This paper aims to offer modification and formalization of the model assessing the difficulty of a language based on the system designed by Foreign Language Institute (FSI) to apply to any language from the viewpoint of English and to test it on well-known features of Chinese including phonetics, lexis, grammar, writing systems, etc. Therefore, modeling is the major method used in this study. The paper presents the structure of the model, its criteria (variables), changes of its output and offers mathematical formalization with further calculations. As a result, two parameters were calculated for Chinese: structure complicatedness value that turned out 69 % and FSI index that was approximately 4,573. These values enable to attribute Chinese to category V. This means that, being an average native English speaker, an average learner of Chinese as L2, will need 85 weeks to achieve level B2/C1.

Keywords: English language, difficulty in learning a language, mathematical model, formalization principle, layer of language, language similarity.

For citation: Osadchi A. A., Androsova S. V. The model to assess the difficulty of a language: Chinese from the view point of English. Eurasian Humanitarian Journal. 2025;1:21-29. (In Russ.).

Введение и актуальность

Данная статья посвящена попытке формализации существующей модели оценки сложности английского языка. Модель основана на типологическом подходе классификации языков, выработанном на опыте Отраслевой лаборатории машинного перевода Российского государственного педагогического университета им. А. И. Герцена (Коваль, Пиотровский и др., 1993; Koval, Van Nunen 1999; Коваль, Ван Нюнэн 1999; Koval et al. 2000) [Цит. по: Коваль 2005: 6–7]], и ориентирована на педагогов и лингвистов с целью разработки курсов для студентов, изучающих второй язык как иностранный [Jackson, Kaplan 1999: 72], а также на полевых лингвистов и миссионеров. Методология может быть применена и для создания более разветвлённой модели оценки сложности языков с точки зрения русского и других языков [Климцев, Каллаур 2022; Cruttenden 2001; Dan 2021].

Математическая модель является приближенной, и не отражает индивидуальных особенностей, но всё же она основана на 70-летнем опыте преподавания широкого спектра языков, накопленном Институтом дипломатической службы (Foreign Service Institute). FSI создал классификацию языков в зависимости от среднего времени, которое требуется студентам для достижения «профессионального рабочего уровня» B2/C1. Эта система является наиболее популярной, что обуславливает её выбор в данной научной статье.

Входными параметрами модели являются лингвистические факторы того или иного языка (фонетические, грамматические и лексические особенности, графика и орфография и т. п.), а также экстралингвистические факторы (наличие ресурсов для изучения), выходным

же параметром является примерное время, которое нужно на изучение языка. Система FSI состоит из 5 категорий сложности языков, требующих для достижения уровня B2/C1 как минимум 24, 30, 36, 44 и 88 недель соответственно.

Основная часть

Постановка задачи

Разработчики системы FSI выделяют следующие факторы, которые позволяют присвоить языку определенную категорию: грамматика, сходство вокабуляра, фонетика, письменность, языковое родство, диалекты, ресурсы для изучения [FSI Language Courses 2024].

Для формализации модели оттолкнемся от того, 1) какие критерии (переменные) представляют наибольшую значимость, 2) как ведет себя функция выходных данных – сколько недель нужно для изучения языка той или иной категории.

Прогрессия времени изучения языка в зависимости от категории увеличивается неравномерно. Кроме того, более сложные элементы языка сильнее влияют на сложность языка в целом. При составлении функций переменных предлагается использовать функцию, имеющую следующий вид:

$$y = \sqrt[z]{x}, \tag{1}$$

где x – некоторая переменная, а значение z зависит от важности параметра. График данной функции будет образовывать горизонтальную параболу, ограниченную $[x, y] \geq 0$ при условии, если z – четное число. То есть, мы не будем брать $z \geq 3$. Ось x означает уровень сложности языка, а ось y – время изучения языка.

Параметры вычисляются в порядке возрастания важности для создания порядка арифметических действий. Поскольку мы имеем дело с усредненной оценкой, сложим эти параметры и вычислим среднее арифметическое для каждой скобки.

Итоговая формула имеет следующий вид:

$$LD = \frac{\frac{\sqrt{G} + \frac{\sqrt[3]{V_0} + \sqrt[3]{V_1} + \sqrt[3]{D}}{3} + \sqrt{P_1} + \sqrt[3]{P_0}}{4} + \sqrt[3]{W}}{2} \tag{2}$$

где LD (Language Difficulty) – уровень структурной сложности языка. Для близкородственных языков формула имеет следующий вид:

$$LD = \frac{\frac{\sqrt{G} + \frac{\sqrt[3]{V_0} + \sqrt[3]{V_1} + \sqrt[3]{D}}{3} + \sqrt{P_1} + \sqrt[3]{P_0}}{4} + \sqrt[3]{\frac{W \cdot D}{2}}}{2} \tag{3}$$

Отдельным случаем являются бесписьменные языки ($W = 0$). Для бесписьменных языков формула имеет следующий вид:

$$LD = \frac{\sqrt{G} + \frac{\sqrt[3]{Va} + \sqrt[3]{Vb} + \sqrt[3]{D}}{3} + \sqrt{Ph} + \sqrt[3]{Po}}{4} + \sqrt[3]{I} ; \quad (4)$$

Необходимо конвертировать получившиеся значения в шкалу FSI:

$$FSI = LD * 2,639; \quad (5)$$

Для языков с $D < 2,636$ значение LD конвертируется по следующей формуле:

$$FSI = LD^2; \quad (6)$$

Наконец, осталось вычислить время, требуемое на изучение языка для овладения уровнем B2/C1. Если $FSI > 2,636$; то количество недель вычисляется следующим образом:

$$T = \frac{100}{5,75 - FSI}; \quad (7)$$

Для $FSI \leq 2,636$:

$$x = 7y + 17,314. \quad (8)$$

Хорда (8) достигает категории I в точке (24,441; 1); категории II в (30,194; 1,816); категории III в (35,975; 2,636). Хорда и парабола функции (7) пересекаются в точке (40,537; 3,283).

Реализация модели оценки сложности языков

Для создания программы было решено воспользоваться возможностями пакета программных продуктов (ППП) Microsoft Office, а именно табличным процессором Microsoft Excel, который предоставляет пользователю широкие возможности в области анализа данных, моделирования, прогнозирования, выполнения различных расчетов и даже написания программ для автоматизации этих расчетов с помощью языка программирования Visual Basic (VBA).

Несмотря на наличие других ППП, позволяющих создавать прикладные программы с более удобным интерфейсом, широкую популярность Excel обрел не только за свои возможности, но и за простоту использования. Более того, Microsoft унифицировала интерфейс офисных приложений (Word, Excel, PowerPoint), поэтому, освоив одну из программ, некоторые инструменты в остальных программах пакета будут знакомыми [MSoffice-Prowork.com 2024].

Итак, параметры модели в порядке убывания важности расположим следующим образом: письменность – W (Writing); родство языка – D (Distance); сходство вокабуляра – Vo (Vocabulary); грамматика – G (Grammar); фонетика – Ph (Phonetics); разновидности языка – Va (Varieties); ресурсы для изучения – R (Resources). Все параметры в модели являются вещественными числами (числами с плавающей запятой), и принимают значения от 1 до 5 (кроме сложности письма W, который может принимать значение 0, об этом ниже).

Формальные критерии оценки языка приведены в комментариях к ячейкам (рис. 1). Прикладная программа представляет собой таблицу, к которой можно добавлять новые колонки для каждого нового языка (рис. 1). Для описания каждого параметра представлены комментарии, где даются формальные критерии для оценивания параметров языка. Общий вид блок-схемы алгоритма выполнения программы представлен на рисунке 2.

Calculated results for natlangs and well-known conlangs						
Your Language	Dutch	Spanish	Italian	French	Norwegian	
Grammar (G)			2	2	1	1
Vocabulary (Vo)			2	2	2	1
Varieties (Va)			3	4,5	3	4
Distance (D)			1	2	1	1,5
Phonetics (Ph)			2	1	3	2
Popularity (Po)			1	2	2	3
Writing (W)			2	1	3	2
CALC RESULT (LD)			26094047	1,153393636	1,217002896	1,216076462
STRUCT DIFFICULTY			13%	15%	22%	22%
FSI INDEX			1,268	1,33	1,481	1,479
RATING			I	I	I	I
WEEKS TO LEARN (T)			26	27	28	28

Рисунок 1. Интерфейс прикладной программы

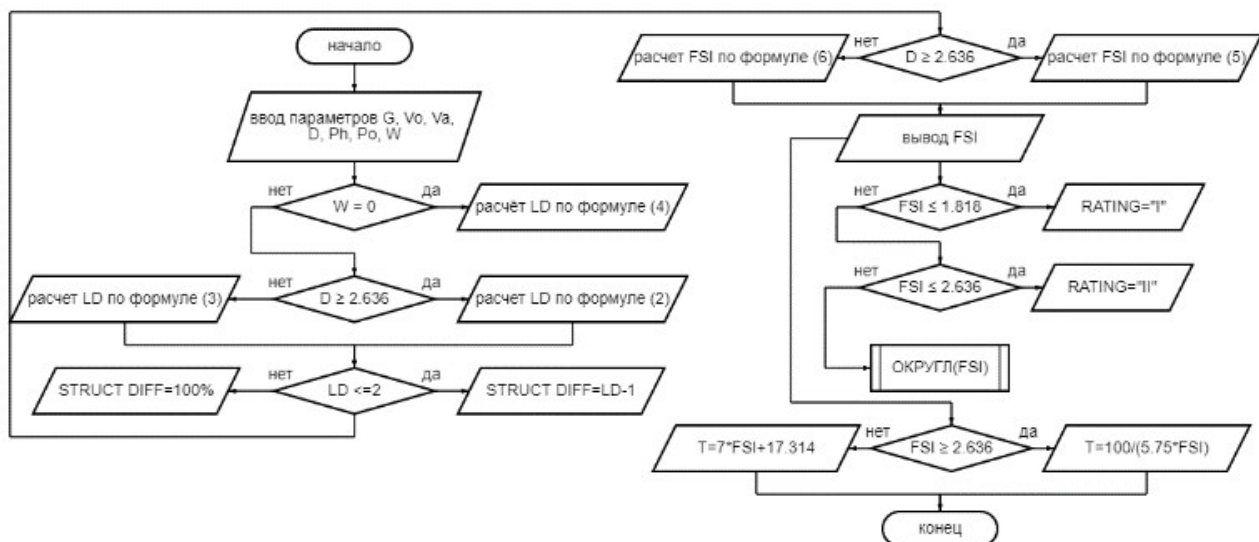


Рисунок 2. Блок-схема выполнения программы

Вычислительный эксперимент

Проведем оценку сложности китайского языка с точки зрения английского. FSI относит язык к категории V [FSI Language Courses 2024]. Установим параметры, исходя из следующих формальных критериев:

Грамматика (G): китайский относится к сино-тибетской языковой семье, является изолирующим языком со строгим порядком слов SVO. **3 балла.**

Сходство вокабуляра (Vo): В китайском языке от 100 до 350 тыс. слов, 3426 из них (1–3 %) заимствованы из английского, что дает оценку в 4 балла. Однако, в силу слоговой природы китайского языка, иностранные слова сильно искажаются при передаче на китайский [Zhou 2015: 60], что повышает оценку параметра до **5 баллов.**

Диалекты и социолекты (Va): большая диалектная раздробленность, носители китайского с севера не способны понимать носителей с юга. **5 баллов.**

Степень языкового родства (D): китайский относится к сино-тибетским языкам, контакты с индоевропейскими не повлияли на структуру языка. **5 баллов.**

Фонетика (Ph): некоторые ряды согласных звуков похожи друг на друга, например, *zh* и *ch*, *c* и *z*, *s* и *x*, и т. д. Четыре тона гласных способны менять смысл слова [Задоевко 2010]. **4 балла.**

Популярность (Po): китайский является одним из самых распространенных языков мира, уступая разве что английскому. **1 балл.**

Письменность: существует более 50 тыс. иероглифов, из которых для теста HSK 3-го уровня требуется освоить 600 иероглифов, а для HSK 6-го уровня – 2600. **5 баллов.**

Итого: значение структурной сложности составляет 69 %; индекс FSI $\approx 4,573$; категория V; 85 недель для достижения уровня B2/C1.

Заключение

В статье рассмотрена методика обучения курсов FSI, по которой были даны эти оценки, что безусловно пригодится в дальнейших исследованиях для создания модели оценки сложности языков с точки зрения не только русского, но и других языков. В ходе исследования была создана математическая модель, способная отражать влияние данных параметров на сложность языка, и создана программа на базе табличного процессора Microsoft Excel, позволяющая просто, удобно и адекватно рассчитывать сложность языков.

Расчеты, произведённые для китайского языка с точки зрения английского языка, показали, что значение структурной сложности составляет 69 %; индекс FSI $\approx 4,573$. Это позволяет присвоить китайскому языку категорию V. Данная категория означает, что среднестатистическому человеку, изучающему китайский язык, для которого родным является английский, потребуется 85 недель для достижения уровня B2/C1.

Хотя разработанная модель адекватно сработала относительно оригинальной системы FSI для китайского языка, она нуждается в апробации на других языках для проверки её универсальности.

Список литературы

1. Задоевко Т. П., Хуан Ш. Начальный курс китайского языка. Ч. I. 5-е изд., испр. и доп. М. : Восточная книга, 2010. 304 с.
2. Занятие 1. Представление Excel – MSoffice-Prowork.com. URL: <https://msoffice-prowork.com/courses/excel/excelpro/excelpro-lesson1> (дата обращения: 12.01.2024).
3. Климцев К. Е., Каллаур В. С. Сложности изучения азиатских языков (на примере китайского, корейского, японского языков) // Постулат. 2022. № 4-1 (78). URL: <https://www.e-postulat.ru/index.php/Postulat/article/view/4179/4236> (дата обращения: 10.01.2024). EDN: SKXAEF
4. Коваль С. А. Лингвистические проблемы компьютерной морфологии. СПб. : СПбГУ, 2005. 151 с. ISBN: 5-288-03731-0 EDN: QSBVCL
5. Cruttenden A. Gimson's Pronunciation of English. Beijing: Foreign Language Teaching and Research Press, 2001. 394 p.
6. Dan D., Wei Ch. A Study of Modern Chinese Typological Characteristics from the Perspectives of Linguistic Typology // Theory and Practice in Language Studies. Vol. 11, No. 7. 2021 P. 853–857. DOI: 10.17507/tpls.1107.12 EDN: LMKWBB
7. FSI language difficulty. URL: <https://www.fsi-language-courses.org/blog/fsi-language-difficulty/> (дата обращения: 08.01.2024).
8. Jackson F. H., Kaplan M. A. Lessons learned from fifty years of theory and practice in government language teaching // Georgetown University round table on languages and linguistics, 1999. Washington, DC: Georetown University Press, 1999. P. 71–87. URL: http://sealang.net/archives/sla/gurt_1999_07.pdf (дата обращения: 10.01.2024).
9. Zhou A. R. Hybridity of Hanyu: Classification and Characteristics of Loanwords in Mandarin Chinese // 常葉大学経営学部紀要 第3巻第1号, 2015年9月, 57-67頁 [Вестник факультета экономики и менеджмента Университета Цунейба. 2015. Т. 3, № 1. С. 57-67.]. URL: <https://tokoha-u.repo.nii.ac.jp/record/647/files/3-1%E3%83%81%E3%83%A7%E3%82%A6.pdf> (дата обращения: 12.01.2024).
10. 杨琼, 基于类比语料库的汉文训读词汇特征研究。-北京:外语教学与研究出版社, 2023年, 第55卷第4期。 [Ян Ц. Исследование лексических особенностей обучения чтению на китайском языке на основе корпуса. Пекин : Изд-во преподавания и исследования иностранных языков. 2023. Vol. 55, No. 4. P. 532-543].

References

1. Zadoenko T. P., Khuan Sh. Nachal'nyy kurs kitayskogo yazyka [Beginner's Course of Chinese Language]. Part I. 5th ed., Moscow, Vostochnaya kniga, 2010, 304 p. (In Russ.).
2. Zanyatie 1. Predstavlenie Excel - MSoffice-Prowork.com [Lesson 1. Excel Presentation - MSoffice-Prowork.com]. (In Russ.). Available at: <https://msoffice-prowork.com/courses/excel/excelpro/excelpro-lesson1> (accessed: 12.01.2024).
3. Klimtsev K. E., Kallaur V. S. Slozhnosti izucheniya aziatskikh yazykov (na primere kitayskogo, koreyskogo, yaponskogo yazykov) [Difficulties in learning Asian languages (using Chinese, Korean, Japanese as an example)]. *Postulat* [Postulate]. 2022, no. 4-1 (78). (In Russ.). Available at: <https://www.e-postulat.ru/index.php/Postulat/article/view/4179/4236> (accessed: 10.01.2024). EDN: SKXAEF (In Russ.).
4. Koval' S. A. Lingvisticheskie problemy kompyuternoy morfologii [Linguistic problems of computer morphology]. Saint Petersburg, SPbGU, 2005, 151 p. ISBN: 5-288-03731-0 EDN: QSBVCL (In Russ.).
5. Cruttenden A. Gimson's Pronunciation of English. Beijing: Foreign Language Teaching and Research Press, 2001. 394 p.
6. Dan D., Wei Ch. A Study of Modern Chinese Typological Characteristics from the Perspectives of Linguistic Typology // *Theory and Practice in Language Studies*. Vol. 11, No. 7. 2021 P. 853–857. DOI: 10.17507/tpls.1107.12 EDN: LMKWBB
7. FSI language difficulty. URL: <https://www.fsi-language-courses.org/blog/fsi-language-difficulty/> (дата обращения: 08.01.2024).
8. Jackson F. H., Kaplan M. A. Lessons learned from fifty years of theory and practice in government language teaching // *Georgetown University round table on languages and linguistics*, 1999. Washington, DC: Georetown University Press, 1999. P. 71–87. URL: http://sealang.net/archives/sla/gurt_1999_07.pdf (дата обращения: 10.01.2024).
9. Chzhou A. R. Gibridnost' han'juj: klassifikacija i harakteristiki zaimstvovaniy v kitajskom jazyke [Hybridity of Hanyu: Classification and Characteristics of Loanwords in Mandarin Chinese]. 常葉大学経営学部紀要 第3巻第1号, 2015年9月, 57-67頁. *Vestnik fakul'teta ekonomiki i menedzhmenta Universiteta Tsuneyba* [Bulletin of the Faculty of Economics and Management, Tsuneova University]. 2015, vol. 3, no. 1, pp. 57-67.]. (In Chinese). Available at: <https://tokoha-u.repo.nii.ac.jp/record/647/files/3-1%E3%83%81%E3%83%A7%E3%82%A6.pdf> (accessed: 12.01.2024).
10. Jan C. Issledovanie leksicheskikh osobennostej obuchenija chteniju na kitajskom jazyke na osnove korpusa [A Corpus-Based Study on Lexical Features of Chinese Reading Teaching]. 杨琼, 基于类比语料库的汉文训读词汇特征研究. -北京:外语教学与研究出版社, 2023年, 第55卷第4期. Pekin, Izdatel'stvo prepodavaniya i issledovaniya inostrannyh jazykov, 2023, Vol. 55, no. 4, pp. 532-543. (In Chinese).

Информация об авторах

А. А. Осадчий – магистрант, кафедра иностранных языков,
Амурский государственный университет;

С. В. Андросова – доктор филологических наук, профессор,
кафедра иностранных языков, Амурский государственный университет.

Information about the authors

A. A. Osadchiy – Master Student, Amur State University;

S. V. Androsova – Grand Ph. D. (Philology), Professor,
Foreign Languages Department, Amur State University.

Статья поступила в редакцию 12.01.2025; одобрена после рецензирования 10.02.2025;
принята к публикации 20.02.2025.

The article was submitted 12.01.2025; approved after reviewing 10.02.2025; accepted for
publication 20.02.2025.